



```

206 oSerThrThrAlaSerLysIleSerAlaLeuGluGlyLeuLeuLysProG 223
|||||
661 ATCCACTACAGCTCAAAATAATCTCAGCTCTGGAGGGCTGCTCAAGCCCC 612
|||||
223 InThrPro. SerTyAsnHisThrArgLeuHisArgGlnArgAlaLe 239
|||||
611 AGAGCCAGGCTACACACACACACAGGCGTCCAGCAGCAGAGCACT 562
|||||
239 uAspTyArgSerClnSerGlyArgGluGlyGlnGlyPheHisIleLeu 256
|||||
561 GAGCATGGCTTCAGCTCTGGAGGGAAGGCCAAGATTTCATCTCTGA 512
|||||
256 leProThrIleLeuGlyLeuPheLeuAlaLeuLeuGlyLeuValVal 272
|||||
511 TCCCGACCATCTGGGCGCTTTCTGTGTGGCACTTCTGGGCTGGTG 462
|||||
273 LysArgAlaValGluArgArgLysAlaLeuSerArgAlaArgLe 289
|||||
461 AAAGGGCGCTGAAGAGGAGAAAGCCCTCTCCAGCGGGCCCGCACT 412
|||||
289 uAlaValArgMetArgAlaLeuGluSerSerGlnArgProArgGlySerP 306
|||||
411 GGCCTGTAGAGTGGCGCCCTGGAGAGCTCCAGAGGCGCCGCGGGTGC 362
|||||
306 rArgProArgSerGlnAsnAsnIleTySerAlaCysProArgArgAla 322
|||||
361 CGCGACCGCGCTCCAAACACATCTACAGCGCTGCGCGGGCGGCT 312
|||||
323 ArgGlyAlaAspAlaGlyThrGlyAlaProValProGlyProG 339
|||||
311 CGTGGAGCGGCGCTGAGGACAGGAGGAGGCGCCGTTCCCGGCGCGG 262
|||||
339 yAlaProLeuProAlaProLeuGlnValSerGluSerProTrpLeuH 356
|||||
261 AGCGCGCTGGCGCGCGCGCTGCAGGTGTTGAATCTCCCTGGCTCC 212
|||||
356 iAlaProSerLeuLysThrSerCysGluTyTrValSerLeuTyHisGln 372
|||||
211 ATGCCCCATCTGTGAGACAGCTGTGAATACGTGAGCTCTACACACAG 162
|||||
373 ProAlaAlaMetMetGluAspSerAspSerAspTyTrIleAsnValPr 389
|||||
161 CCTGGCGCATGATGGAGGACAGTTCAGATGACTACATCAATGTTC 112
|||||
389 oAla 390
|||||
111 TGCC 108

seq_name: gb_est7:AA456707

seq_documentation_block:
LOCUS AA456707 404 bp mRNA EST 06-JUN-1997
DEFINITION aa13904.r1 Soares_NhMPu_S1 Homo sapiens cDNA clone IMAGE:813174
5', mRNA sequence.
ACCESSION AA456707
VERSION AA456707.1 GI:2179283
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 404)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,
Moore,B., Scheinberg,K., Steptoe,M., Tan,F., Theising,B.,
White,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE WashU-Merck EST project 1997
JOURNAL Unpublished (1997)
COMMENT On Jan 14, 1998 this sequence version replaced gi:1877745.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

```

Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available royalty-free through LNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: -28m13 rev2 ET from Amersham  
 High quality sequence stop: 390.

# FEATURES

```

1..404
Location/Qualifiers
Source
/organism="Homo sapiens"
/db_xref="GB:6044084"
/db_xref="taxon:9606"
/clone="IMAGE:813174"
/clone_lib="Soares_NhMPu_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/note="Organ: mixed (see below); Vector: pMT3D-Pac
(Pharmacia) with a modified polylinker; Site.1: Not I;
Site.2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NbHM, pregnant uterus
NbHPU, and fetal heart NbH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."
BASE COUNT 71 a 153 c 114 g 66 t
ORIGIN

```

alignment\_scores:  
 Quality: 551.00 Length: 111  
 Ratio: 5.009 Gaps: 0  
 Percent Similarity: 99.099 Percent Identity: 98.198

# alignment\_block:

```

US-09-135-238B-2 x AA456707 ..
Align seg 1/1 to: AA456707 from: 1 to: 404

280 LysAlaLeuSerArgArgAlaArgLeuAlaValArgMetArgAlaLe 296
|||||
1 GAAGCGCTCTCCAGCGCGCGCGCGCTGCGGTGAGGATGCGCGCCT 50
|||||
296 uGluSerSerGlnArgProArgGlySerProArgProArgSerGlnAsnA 313
|||||
51 GGAGAGCTCCAGAGCGCGCGCGGTGCGCGGA. CCGCGCTCCCAACA 99
|||||
313 snIleTySerAlaCysProArgArgAlaArgGlyAlaAspAlaGly 329
|||||
100 ACATCTACAGCGCTCCCGCGCGCGCTCGTGAGCGGAGCGTGCAGGC 149
|||||
330 ThrGlyGluAlaProValProGlyProGlyAlaProLeuProProAlaPr 346
|||||
150 ACAGGGGAGGCGCGCTTCCCGCGCGGAGCGCGTTCGCCCGCCGCC 199
|||||
346 oLeuGlnValSerGluSerProTrpLeuHisAlaProSerLeuLysThrS 363
|||||
200 GGTGAGGTGTCTGAATCTCCCTGGCTCCATGCCCTCTCTGAGACCA 249
|||||
363 eCysGlyTyTrValSerLeuTyTrHisGlnProAlaAlaMetMetGluAsp 379
|||||
250 GCTGTGAATACGTGAGCGCTTACCACAGCGCTGCCGCCATGATGGAGGAC 299
|||||
380 SerAspSerAspTyTrIleAsnValProAla 390
|||||
300 AGTGATTCAGATGACTACATCAATGTTCTCTGCC 332

```

seq\_name: gb\_est42:AW402953

seq\_documentation\_block:







272 allYsArgAlaValGluArgArgLysAlaLeuSerArgAlaArgArg 288  
 |||||  
 550 TAAAGACCCATTCAAGAGGAGAGAGCTCTCCAGACGTCGGGCGCA 599  
 |||||  
 289 LeuAlaValArgMetArg 294  
 |||||  
 600 CTGGCGATGAGGAGCGCA 617

seq\_name: gb\_est45:AW655218

seq\_documentation\_block:  
 LOCUS AW655218 494 bp mRNA EST 05-APR-2000  
 DEFINITION 105873 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.  
 ACCESSION AW655218  
 VERSION AW655218.1 GI:7421044  
 KEYWORDS EST.  
 SOURCE Bos taurus.  
 ORGANISM Bos taurus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
 Bovidae; Bovinae; Bos.  
 1 (bases 1 to 494)  
 Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,  
 Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A. and  
 Keele,J.W.

TITLE Design and use of four pooled tissue normalized cDNA libraries for EST discovery in cattle

JOURNAL Unpublished (2000)  
 COMMENT On Mar 8, 1999 this sequence version replaced gi:4388128.  
 Contact: Smith TPL  
 USDA, ARS, US Meat Animal Research Center  
 PO Box 166, Clay Center, NE 68933-0166, USA  
 Tel: 402 762 4366  
 Fax: 402 762 4390  
 Email: smith@mail.marc.usda.gov  
 Single pass sequencing. Bases called and alt trimmed with phred  
 v0.980904.e. Vector identified by cross\_match with the -minscore 18  
 and -minmatch 12 options.

PCR Primers  
 FORWARD: AGGAACACCTATGACGAT  
 BACKWARD: GTTTCGCCAGTCAGCAGC  
 Plate: 86 row: M column: 8  
 Seq primer: ATTTAGTGACACTATAG.

FEATURES  
 source

1..494  
 /organism="Bos taurus"  
 /db\_xref="taxon:9913"  
 /clone\_lib="MARC 1BOV"  
 /tissue\_type="pooled"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV SPORT6; Site\_1: XbaI; Site\_2: XhoI;  
 Library made from pooled tissue from lymph node, ovary,  
 fat, hypothalamus, and pituitary."

BASE COUNT  
 ORIGIN

118 a 125 c 150 g 101 t

alignment\_scores:

Quality: 455.00 Length: 128  
 Ratio: 4.213 Gaps: 0  
 Percent Similarity: 84.375 Percent Identity: 64.844

alignment\_block:

US-09-135-238B-2 x AW655218

Align seg 1/1 to: AW655218 from: 1 to: 494

1 MetAspArgTrpLeuTrpProLeuTyrrPheLeuProValSerGlyAlaLe 17  
 |||||  
 111 ATGGACCTGTGGCTTTGGGCGCTTTACTTCTCCAGTACTGCGGAGCCCC 160  
 |||||  
 17 uArgIleLeuProGluValLysValGluGlyGluLeuGlySerValr 34

161 GAAGGTCCTTCCAGAAATAAGATGGAAGGAATGCTGGGAGGATCCATT 210  
 |||||  
 34 hrIleLysCysProLeuProGluMetHisValArgIleTyrrLeuCysArg 50  
 |||||  
 211 CCATCGAGTCCCACTTCTCGAAACGCGATGTGAGGATATATCTGTGCCG 260  
 |||||  
 51 GluMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsnPh 67  
 |||||  
 261 ACCATTGACGAGTCTGGAAGATGCACCCGCTGTGTCCAGCAACAATA 310  
 |||||  
 67 eIleLysAlaGluTyrrLysGlyArgValThrLeuLysGlnTyrrProArgL 84  
 |||||  
 311 CGTAAGAGAGAAATCAAGACCGGACGCTGGGACGATGTCCGGACA 360  
 |||||  
 84 yAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerGly 100  
 |||||  
 361 GGAATCTGTTCTTGTGTGATGACAGAGCTGACCAAGAATGACAGCGG 410  
 |||||  
 101 ValTyrrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGln 117  
 |||||  
 411 ATCTATGCTGTGGGTGGCGGGAACACAGACCGAGGAGGAGCCATCA 460  
 |||||  
 117 sValThrLeuAsnValHisSerGluTyrrGluPro 128  
 |||||  
 461 GATCACCTGACTGTTTACAGCGCTTACTAGCCA 494

seq\_name: gb\_est42:AW444691

seq\_documentation\_block:  
 LOCUS AW444691 482 bp mRNA EST 17-FEB-2000  
 DEFINITION UI-H-B13-ajw-g-12-0-UI.s1 NCI\_CGAP\_Sub5 Homo sapiens cDNA clone  
 IMAGE:2733502 3', mRNA sequence.  
 ACCESSION AW444691  
 VERSION AW444691.1 GI:5986453  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 482)  
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)

JOURNAL  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 The sequence contained an oligo-dT track that was present in the  
 oligonucleotide that was used to prime the synthesis of first  
 strand cDNA and therefore this may represent a bonafide poly A  
 tail. cDNA library Preparation: M.B. Soares Lab Clone distribution:  
 NCI-CGAP clone distribution information can be found through the  
 I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
 The following repetitive  
 elements were found in this cDNA sequence: 1-30,  
 >POLY\_A#Simple\_repeat  
 Seq primer: M13 Forward  
 POLYA=Yes.

FEATURES

source

Location/Qualifiers  
 1..482  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="NCI\_CGAP\_Sub5"  
 /lab\_host="DH10B (Life Technologies)"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: Not I; Site\_2: Eco RI; NCI\_CGAP\_Sub5  
 is a subtracted library derived from NCI\_CGAP\_Sub4. The  
 NCI\_CGAP\_Sub5 library had 3 million recombinants. A  
 single-stranded DNA preparation of NCI\_CGAP\_Sub4 was used  
 as a tracer in a subtractive hybridization with a driver  
 comprising: the IMAGE pool (NCI\_CGAP\_Kid3 pool 1 LLAM

3334-3337, 3682-3683, 3798-3803 (IMAGE Clones)  
1322376-1323911, 1456008-1456775, 1500552-1502855);  
NCI CGAP Kid5 pool 1 LAM 3338-3342, 3722-3725, 3776-3778  
(IMAGE Clones) 1323912-1325831, 1471368-1472903,  
1492104-1493255); NCI CGAP Lu5 pool 1 LAM 3575-3582,  
3851-3854 (IMAGE Clones)  
1414920-1417991, 1520904-1522439); NCI CGAP GC4 pool 1 LAM  
3164-3167, 3716-3720, 3733-3735 (IMAGE Clones)  
1257096-1258631, 1469064-1470983, 1475592-1476743);  
NCI CGAP Pr22 pool 1 LAM 2457-2459, 2758-2759, 3062-3068  
(IMAGE Clones) 985608-986759, 1101192-1101959,  
1217928-1220615); NCI CGAP Co10 pool 1 LAM 2644-2653,  
2871-2872 (IMAGE Clones)  
1057416-1061255, 1144584-1145351). (10% of the driver  
population), plus a pool of 3,840 arrayed clones from  
NCI CGAP Sub1 (IMAGE Clones) 2708616-2710535) and  
NCI CGAP Sub2 (IMAGE Clones) 2710536-2712455) (10% of  
the driver population), plus a pool of 11,136 clones from  
NCI CGAP Sub3 (IMAGE Clones) 2712456-2723591) (10% of the  
driver population), plus a pool of 5,472 clones from  
NCI CGAP Sub4 (IMAGE Clones) 2723592-2728969) (70% of the  
driver population). Subtraction was performed as  
previously described [Bonaldi, Lennon & Soares (1996):  
Normalization and Subtraction: Two Approaches To  
Facilitate Gene Discovery. Genome Research 6, 791-806. ;  
TAG\_LIB=NCI CGAP\_Kid5; TAG\_TISSUE=kidney; TAG\_SEQ=ATTC";  
BASE COUNT 103 a 107 c 111 g 161 t  
ORIGIN

alignment\_scores:  
Quality: 444.50 Length: 117  
Ratio: 4.401 Gaps: 3  
Percent Similarity: 86.325 Percent Identity: 80.342

alignment\_block:

US-09-135-238B-2 x AW444691/rev ..

Align seg 1/1 to reverse of: AW444691 from: 1 to: 482

41 GluMetHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyTh 57  
|||||  
481 GAAATGATCTGAGGATATCTGCGGGAGATGCTGATCGAAC 432  
57 rCysGlyThrValSerThrThrAsnPhelIleLysAlaGluTyrLysG 74  
|||||  
431 ATGTGGTACCGTGGTATCCACCACCACTTCATCAAGCAGATACAAG 382  
74 lYArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGlu 90  
|||||  
381 GCCAGTTACTCTGAAGCAATACCCAGCAAGAACTCTCTCTAGTGGAG 332  
91 ValThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaG 107  
|||||  
331 GTAAACAGCTGACAGAAAGTGACAGCGAGTCTATGCTCGGGAGCGG 282  
107 yMetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnValHis 124  
|||||  
281 CATGACACAGACCGGGAAAGACCCAGAAAGTCAACCTGAATGTCACA 232  
124 er...GluTyrGluProSerTrpGluGlnProMetProGluThrPro 139  
|| : : : : :  
231 CTGGTAGTTCCTCCGCTGATGGAGGCTCAG.....ACTACCCAG 191  
140 LysTrpPheHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSerSe 156  
|||||  
190 AAAATGTTTCACTTGGGAGATTG...GAACAGCCATCTTAACGTGAACAG 144  
156 r 156  
143 T 143  
seq\_name: gb\_est41:AW327067

seq\_documentation\_block:  
LOCUS AW327067 359 bp mRNA EST 27-JAN-2000  
DEFINITION 20516 MARC 1B0V Bos taurus cDNA 5', mRNA sequence.  
ACCESSION AW327067  
VERSION AW327067.1 GI:6762988  
SOURCE EST.  
ORGANISM Bos taurus.  
REFERENCE 1 (bases 1 to 359)  
Smith, R.P.L., Casas, E., Stone, R.T., Heaton, M.P., Grosse, W.M.,  
Bennett, G.A., Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A. and  
Keefe, J.W.  
TITLE Design and use of four pooled tissue normalized cDNA libraries for  
EST discovery in cattle  
JOURNAL Unpublished (2000)  
COMMENT On Jan 6, 2000 this sequence version replaced gi:6676592.  
Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smithemail.marc.usda.gov  
Single pass sequencing. Bases called and trimmed with phred  
v0.980904.e. Vector identified by cross\_match with the -minscore 20  
and -minmatch 12 options.  
PCR Primers  
FORWARD: AGGAACAGCTATGACCAT  
BACKWARD: GTTTCCTCAGTCACGAGC  
Plate: 13 row: K column: 5  
Seq primer: ATTAGGTGACACTATAG.  
FEATURES  
source  
1..359  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_lib="MARC 1B0V"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/note="vector: pCMV SPORT6; Site\_1: XbaI; Site\_2: XhoI;  
Library made from pooled tissue from lymph node, ovary,  
fat, hypothalamus, and pituitary."  
BASE COUNT 90 a 111 c 87 g 71 t  
ORIGIN  
alignment\_scores:  
Quality: 415.00 Length: 119  
Ratio: 4.192 Gaps: 1  
Percent Similarity: 83.193 Percent Identity: 64.706  
alignment\_block:  
US-09-135-238B-2 x AW327067 ..  
Align seg 1/1 to: AW327067 from: 1 to: 359  
76 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 92  
|||||  
2 GTCACTGGAGCAGTCTCGGACAGAACTCTGCTTCTGTGTGATGAC 51  
92 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 109  
|||||  
52 AGAGCTGACCAAGAAATCACAGCGGATCTATGCTGTGGGGTGGCGGGA 101  
109 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 125  
|||||  
102 ACACAGCCGAGGCAAGACCCAGACATCACCTGACTGTTTCACAGCGTT 151  
126 TyrGluProSerTrpGluGlnProMetProGluThrProLysTrpPh 142  
|||||  
152 TACGAGCCATCTCGGGAAGAGCGATGCTGCTCCAGCATGGTT 201



and Eco RI sites of the modified p7T3 vector. RNA provided by Dr. Bertrand Jordan. Library went through three rounds of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 108 a 158 c 117 g 77 t  
ORIGIN

alignment\_scores:  
Quality: 360.00 Length: 153  
Ratio: 3.130 Gaps: 5  
Percent Similarity: 75.163 Percent Identity: 54.902

alignment\_block:  
US-09-135-238B-2 x AA174968 ..

Align seg 1/1 to: AA174968 from: 1 to: 460

```

188 ProArgValSerArgAlaSerSerValAlaGlyAspLysProArgThrPh 204
7 CCCAGAGTTTACAGAGCATTTCTGTGTCAGCTACCAAGTCCCGAGCGCT 56
204 eLeuProSerThrThrAlaSerLysLysSerAlaLeuGlyLeuLeuL 221
57 CTTGCCAGCAACCCAGCCTCAGAGACTTCCACTCAGCAAGCA...ATCA 103
221 ysProGlnThrProSerTyrAsnHisHisThrArgLeuHisArgGlnArg 237
104 GGGCCCTTAGAGCCAGCTACAGCCACCACCCAGACTTCATGAGCAAGG 153
238 AlaLeuAspTyrGlySerGlnSerGlyArgGlyGlnGly..... 251
154 ACAGCGCCAGCATGGCCACACACTATGGGAGAGAGACCGAGGCTTCACAT 203
252 .....PheHisLeuLeuProThrIleLeuGlyLeuPheL 264
204 CCCCATCCCAAGATTCACATCTGATCCGACCTCTCTGGGCTTCTCT 253
264 eLeuLeuLeuLeuValValLysAlaValAlaValGlnArgLys 280
254 TGCTGGTCTTTGGGACTGTGTAAAGAGCCATCAAGAGGAGGA 303
281 AlaLeuSerArgAlaAlaArgArgLeuAlaValArgMetArgAlaLeuG 297
304 GCTTCTCCAGACGTGGCGGCGGACTAGCATGAGAGCGAGCGCGGG 353
297 userSerGln.Arg.ProArgGlySerPro.....ArgProAr 309
354 GCTTCTCCCGCGCTTCCACACAGCGCGGATGCTCGCAGAGCGCG 403
309 gSerGlnAsnAsnIleTyrSerAlaCysProArgArgAlaArgGlyAlaA 326
404 CTCGACAGACACAGCTCTACAGCGCTGCCCGCGCGGCGGAGCCAG 453
326 SpAla 327
454 ACAGC 458

```

seq\_name: gb\_est4:AA289667

seq\_documentation\_block: 417 bp mRNA EST 14-APR-1997  
LOCUS AA289667  
DEFINITION vb20c07.r1 Soares mouse 3NbMS Mus musculus cDNA clone IMAGE:749484  
5', mRNA sequence.  
ACCESSION AA289667  
VERSION AA289667.1 GI:1937343  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 417)  
AUTHORS Marra.M., Hillier.L., Allen.M., Bowles.M., Dietrich.N., Dubuque.T.,

Geisel.S., Kucaba.T., Lacy.M., Le.M., Martin.J., Morris.M., Schellenberg.K., Steptoe.M., Tan.F., Underwood.K., Moore.B., Theising.B., Wyllie.T., Lennon.G., Soares.B., Wilson.R. and Waterston.R.

TITLE  
JOURNAL  
COMMENT

The WashU-HHMI Mouse EST Project  
Contact: Marra M/Mouse EST Project  
WashU-HHMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: mouseest@wustl.edu

This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:458468

Seq primer: -28ml3 rev2 ET from Amersham  
High quality sequence stop: 405.

FEATURES  
source

```

1..417
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:749484"
/clone_lib="Soares mouse 3NbMS"
/sex="male"
/tissue_type="Spleen"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTACCAATCTCAAGTGGAGCGCGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified p7T3 vector. RNA
provided by Dr. Bertrand Jordan. Library went through
three rounds of normalization, and was constructed by
Bento Soares and M. Fatima Bonaldo."
BASE COUNT 109 a 140 c 91 g 77 t
ORIGIN

```

alignment\_scores:  
Quality: 349.00 Length: 138  
Ratio: 3.231 Gaps: 2  
Percent Similarity: 78.261 Percent Identity: 53.623

alignment\_block:

US-09-135-238B-2 x AA289667 ..

Align seg 1/1 to: AA289667 from: 1 to: 417

```

162 ArgValThrThrProAlaGlnArgGlyLysValProValHisHisSe 178
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4 AAAGTTACCCAGCCACCAAGACTCCAAAGACTGAGGCCCCCTCCGGTTCACGCC 53
178 rSerProThrThrGlnIleThrHisArgProArgValSerArgAlaSerS 195
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
54 CTCGAGCATCACTTCAGTAACCCACATCCAGAGTTTACAGAGCATTTT 103
195 erValAlaGlyAspLysProArgThrPheLeuProSerThrThrAlaSer 211
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
104 CTGTGTACAGTACCAAGTCCCGCGGCTCTCGCCAGCAACACAGCGCTCA 153
212 LysIleSerAlaLeuGlyLeuLeuLysProGlnThrProSerTyrAs 228
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
154 AAGATTCACTCAGCAAGCA...ATCAGGCCCTTAGAGCGCCAGCTACAG 200
228 nHisHisThrArgLeuHisArgGlnAlaLeuAspTyrGlySerGlnS 245
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
201 CCACACACAGAGTCTTCATGAGCAAGGACACGCCACCATGGCCACACT 250

```

245 erGlyArgGluGlyGlnGly.....PheHisIle 254  
 ||||| :|||  
 251 ATGGGAGAGACCGAGGCTTCACATCCCATCCAGAAATTCACATC 300  
 ||||| :|||  
 255 LeuIleProThrIleGlyLeuPheLeuLeuAlaLeuLeuGlyLeuVa 271  
 ||||| :|||  
 301 CTGATTCGACGCTTCCTGGGCTTCTCTGCTGGTCTTTGGGACTGGT 350  
 ||||| :|||  
 271 lVallysArgAlaValGluArgGlyAlaLeuSerArgArgAlaArgA 288  
 ||||| :|||  
 351 GGTAAAGAGCCATCAAGAGGAGGAGGCTCTCTCCAGACGTCGGGCC 400  
 ||||| :|||  
 288 rgLeuAlaValarg 292  
 ||||| :|||  
 401 GACTAGCGATGAGG 414

seq\_name: gb\_est21:AI508502

seq\_documentation\_block:

LOCUS AI508502 568 bp mRNA EST 12-MAR-1999  
 DEFINITION vb34f04.y1 Soares mouse lymph node NBMLN Mus musculus cDNA clone  
 IMAGE:750847 5', mRNA sequence.

ACCESSION AI508502

VERSION AI508502.1 GI:4407407

KEYWORDS EST.

SOURCE house mouse.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

AUTHORS

Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,  
 Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,  
 Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,  
 Ritter,E., Kohn,S., Shin,T., Jackson,F., Cardenas,M., McCann,R.,  
 Waterston,R. and Wilson.R.

The WashU-NCI Mouse EST Project 1999

Unpublished (1999)

On Nov 2, 1998 this sequence version replaced gi:3830418.

Contact: Marra M/WashU-NCI Mouse EST Project 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.

MG1:459831

This read is a RESEQUENCE of a previously sequenced mouse clone  
 This read has been verified (found to hit its original self in the  
 correct orientation)

Putative full length read

vector to vector length is 639

Seq primer: -40RP from Gibco

High quality sequence stop: 458.

FEATURES

source

1..568  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone\_image="IMAGE:750847"  
 /clone\_lib="Soares mouse lymph node NBMLN"  
 /sex="male"  
 /tissue\_type="lymph node"  
 /dev\_stage="4 weeks"  
 /lab\_host="DH10B"

/note="Organ: Lymph node; Vector: pT7T3D-Pac (Pharmacia)  
 with a modified polylinker; Site\_1: Not 1; Site\_2: Eco RI;  
 1st strand cDNA was primed with a Not I - oligo(dT) primer  
 [5].  
 TGTACCATCTGAAGTGGGAGCGCGGCTACTTTTTTTTTTTTTTTTTTTT  
 3]; double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pT7T3 vector. RNA

provided by Dr. Bertrand Jordan. Library constructed and  
 normalized by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 109 a 191 c 144 g 124 t  
 ORIGIN

alignment\_scores:

Quality: 331.50 Length: 115

Ratio: 3.900 Gaps: 1

Percent Similarity: 73.913 Percent Identity: 60.000

alignment\_block:

US-09-135-238B-2 x AI508502 ..

Align seg 1/1 to: AI508502 from: 1 to: 568

280 LysAlaLeuSerArgArgAlaArgLeuAlaValArgMetArgAlaLe 296

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

42 GAAGCCTCTCCAGAGTGGGGCCGACTAGCGATGAGGAGCGAGCGCG 91

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

296 uGluSerSer.....GlnArgProArgGlySerProArgP 308

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

92 GGGGGCTTCCCGCCCTTCCCCACACAGCGCGGATGCTCGCAGAGGC 141

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

308 rGArgSerGlnAsnAsnIleTyrSerAlaCysProArgArgAlaArgGly 324

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

142 CGCGCTCGCAGAACACAGCTCTACAGCGCTGCCCGCGCGCACGGGA 191

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

325 AlaAspAlaAlaGlyThrGlyGluAlaProValProGlyProGlyAlaPr 341

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

192 CCAGACAGCTGGGTCCAGCGAGGCTCGCTCTCAGCCGCCAGCCCTC 241

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

341 oLeuProAlaProLeuGlnValSerGluSerProTrpLeuHisAlaP 358

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

242 AGCGTCCCCGCTTCTCCGAGGTACTTGAAGCTCTCTTGGCCCCACCC 291

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

358 rGSerLeuLysThrSerCysGluTyrValSerLeuTyrHisGlnProAla 374

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

292 CATCTCTGAAGATGAGCTGTGAATACGTGAGCTTGGGTACAGCCTGCT 341

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

375 AlaMetGluAlaSerSerAspAspTyrIleAsnValPro 389

||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

342 GTCACCTGGAGAGCCCTGATTCAGATGATTACATCATATTCCT 386

seq\_name: gb\_est12:AA830944

seq\_documentation\_block:

LOCUS AA830944 485 bp mRNA EST 07-APR-1998

DEFINITION oc58b10.sl NCI\_CGAP\_GCB1 Homo sapiens cDNA clone IMAGE:1353883 3',

mRNA sequence.

ACCESSION AA830944

VERSION AA830944.1 GI:2904043

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

COMMENT

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,

Ph.D., Gerald Marti, M.D.

CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima

Bonaldo, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:



www.bio.llnl.gov/bbrp/image/image.html  
 Insert Length: 983 Std Error: 0.00  
 Seq primer: -40m13 fwd. ET from Amersham  
 High quality sequence stop: 473.

## FEATURES

source  
 1. .485  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="NCI\_CGAP\_GCB1"  
 /tissue\_type="germinal center B cell"  
 /lab\_host="DH10B"  
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
 polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA  
 was prepared from human tonsillar cells enriched for  
 germinal center B cells by flow sorting (CD20+, IgD-),  
 provided by Dr. Louis M. Staudt (NCI); Dr. David Allman  
 (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was  
 primed with a Not I - oligo(dT) primer  
 [5'-TGTTACCAATCTGAAGTGGGCGCGCTCATTTTTTTTTTTT-3'  
 ]. Double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pT7T3 vector. Library  
 went through one round of normalization, and was  
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 109 a 84 c 137 g 155 t  
 ORIGIN  
 alignment\_scores:  
 Quality: 305.00 Length: 59  
 Ratio: 5.169 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-135-238B-2 x AA830944/rev ..

Align seg 1/1 to reverse of: AA830944 from: 1 to: 485

179 SerProThrThrGlnIleThrHisArgProArgValSerArgAlaSerSe 195  
 |||||  
 485 TCCCCACACCAATCAATCACCACCGCTCGAGTCCAGAGCATCTTC 436  
 |||||  
 195 rValaIagLysProArgThrPheLeuProSerThrAlaSerL 212  
 |||||  
 435 AGTAGCAGGTGACAAAGCCGACCTTCCTGCCATCCACTACAGCCTCAA 386  
 |||||  
 212 yIleSerAlaLeuGluGlyLeuLeuLysProGlnThrProSerTyrAsn 228  
 |||||  
 385 AAATCTCAGCTCTGGAGGGGCTGCTCAAGCCCCACAGCCCGAGTACAAC 336  
 |||||  
 229 HisHisThrArgLeuHisArgGlnArg 237  
 |||||  
 335 CACCACACAGGCTGCACAGGCAGG 309

seq\_name: gb\_est8:AA491123

seq\_documentation\_block:  
 LOCUS AA491123 179 bp mRNA EST 18-AUG-1997  
 DEFINITION aa4904.r1 NCI\_CGAP\_GCB1 Homo sapiens cDNA clone IMAGE:824022 5',  
 mRNA sequence.

ACCESSION AA491123

VERSION AA491123.1 GI:2220296

KEYWORDS EST.

SOURCE human.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 179)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

JOURNAL Unpublished (1997)

## COMMENT

Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert\_Strausberg@nih.gov  
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,  
 Ph.D., Gerald Marti, M.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
 Bonaldo, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -28m13 rev1 ET from Amersham

High quality sequence stop: 116.

## FEATURES

## source

1. .179  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="NCI\_CGAP\_GCB1"  
 /tissue\_type="germinal center B cell"  
 /lab\_host="DH10B"  
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
 polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA  
 was prepared from human tonsillar cells enriched for  
 germinal center B cells by flow sorting (CD20+, IgD-),  
 provided by Dr. Louis M. Staudt (NCI); Dr. David Allman  
 (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was  
 primed with a Not I - oligo(dT) primer  
 [5'-TGTTACCAATCTGAAGTGGGCGCGCTCATTTTTTTTTTTT-3'  
 ]. Double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pT7T3 vector. Library  
 went through one round of normalization, and was  
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 46 a 69 c 34 g 30 t  
 ORIGIN

## alignment\_scores:

Quality: 299.00 Length: 59  
 Ratio: 5.068 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-135-238B-2 x AA491123 ..

Align seg 1/1 to: AA491123 from: 1 to: 179

163 ValThrThrProAlaGlnArgGlyLysValProValHisHisSerSe 179  
 |||||  
 3 GTTACACACAGCTCAAGGGGAGGTCCCTCCAGTTCACCATCTCTC 52  
 |||||  
 179 rProThrThrGlnIleThrHisArgProArgValSerArgAlaSerServ 196  
 |||||  
 53 CCCCACCAACCAATCACCACCGACCTCGAGTGTCCAGAGCATCTTCAG 102  
 |||||  
 196 aIaIaGlyAspLysProArgThrPheLeuProSerThrThrAlaSerLys 212  
 |||||  
 103 TACAGGTGACAAAGCCCCGAGACCTTCCTGCCATCCACTACAGCCCAAAA 152  
 |||||  
 213 ILeSerAlaLeuGluGlyLeuLeuLys 221  
 |||||  
 153 ATCTCAGCTCTGGAGGGGCTGCTCAAG 179

seq\_name: gb\_est40:AW291041

## seq\_documentation\_block:

LOCUS AW291041 561 bp mRNA EST 16-JAN-2000  
 DEFINITION UI-H-B12-adj-h-06-0-UI.s1 NCI\_CGAP\_Sub4 Homo sapiens cDNA clone  
 IMAGE:2724707 3', mRNA sequence.  
 ACCESSION AW291041  
 VERSION AW291041.1 GI:6697677

KEYWORDS  
SOURCE

EST.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 561)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On May 18, 1998 this sequence version replaced gi:3137211.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
The sequence contained an oligo-dT track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. cDNA Library Preparation: M.B. Soares lab Clone distribution:  
NCI-CGAP clone distribution information can be found through the  
I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/bbrp/image/image.html](http://www.bio.llnl.gov/bbrp/image/image.html)  
Seq primer: M13 Forward  
POLYA=Yes

FEATURES  
source

Location/Qualifiers  
1...561  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2724707"  
/clone\_lib="NCI-CGAP\_Sub4"  
/lab\_host="DH10B (Life Technologies)"  
/note="vector: pT73D-Pac (Pharmacia) with a modified  
polylinker. Site.1: Not I; Site.2: Eco RI; The  
NCI-CGAP\_Sub4 library is a subtracted library derived from  
the NCI-CGAP\_Sub2 library which is a subtracted library  
derived from the NCI-CGAP\_Sub1 library, which is a  
subtracted library derived from BI. BI constitutes a  
mixture of 21 normalized or subtracted NCI-CGAP  
libraries: NCI-CGAP\_Co4, NCI-CGAP\_Pr22, NCI-CGAP\_Pr28,  
NCI-CGAP\_Co10, NCI-CGAP\_Co16, NCI-CGAP\_Kid5,  
NCI-CGAP\_Kid12, NCI-CGAP\_Kid3, NCI-CGAP\_Kid11,  
NCI-CGAP\_Lym2, NCI-CGAP\_Br2, NCI-CGAP\_Co8, NCI-CGAP\_CLL1,  
NCI-CGAP\_Le12, NCI-CGAP\_Brn23, NCI-CGAP\_Lu5,  
NCI-CGAP\_Lu24, NCI-CGAP\_Lu19, NCI-CGAP\_GC4, NCI-CGAP\_GC6,  
NCI-CGAP\_Brn25. These 21 libraries were pooled and a  
single-stranded DNA preparation of the resulting mixture  
was used as a tracer in a subtractive hybridization with  
a driver whose composition is detailed below:  
NCI-CGAP\_Kid3 pool 1 : LLAM 3334-3337, 3682-3683,  
3798-3803 (IMAGE Clonoids 1322376-1323911,  
1456008-1456775, 1500552-1502855) NCI-CGAP\_Kid5 pool 1 :  
LLAM 3338-3342, 3722-3725, 3776-3778 (IMAGE Clonoids  
1323912-1325831, 1471368-1472903, 1492104-1493255)  
NCI-CGAP\_Lu5 pool 1 : LLAM 3575-3582, 3851-3854 (IMAGE  
Clonoids 1414920-1417991, 1520904-1522439) NCI-CGAP\_GC4  
pool 1 : LLAM 3164-3167, 3716-3720, 3733-3735 (IMAGE  
Clonoids 1257096-1258631, 1469064-1470983,  
1475592-1476743) NCI-CGAP\_Pr22 pool 1 : LLAM 2457-2459,  
2758-2759, 3082-3088 (IMAGE Clonoids 985608-986759,  
1101192-1101959, 1217928-1220615) NCI-CGAP\_Co10 pool 1 :  
LLAM 2644-2653, 2871-2872 (IMAGE Clonoids 1057416-1061255,  
1144584-1145351) Subtraction was performed as previously  
described [Bonaldo, Lennon & Soares (1996): Normalization  
and Subtraction: Two Approaches To Facilitate Gene  
Discovery. Genome Research 6, 791-806].  
TAG\_LIB=NCI-CGAP\_Co10; TAG\_TISSUE=colon; TAG\_SEQ=AAACG"  
121 a 125 c 120 g 194 t 1 others

BASE COUNT  
ORIGIN

alignment\_scores:  
Quality: 292.50 Length: 91  
Ratio: 3.953 Gaps: 3  
Percent Similarity: 81.319 Percent Identity: 72.527

alignment\_block:  
US-09-135-238B-2 x AW291041/rev ..

Align seg 1/1 to reverse of: AW291041 from: 1 to: 561  
67 PheileLysAlaGluTyrLysGlyArgValThrLeuLysGluTyrProAr 83  
|||||  
561 TTCATCAGCGAGATACAAAGGCCCGAGTTACTCTGAAGCAATACCCAGG 512  
83 gLysAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerG 100  
|||||  
511 CAAGAATCTCTTCTAGTGGAGGTAACACAGCTGACAGAAAGTGACAGCG 462  
100 LyValTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGln 116  
|||||  
461 GAGTCTATGCTCGCGAGCGGGATGACACACACACCGGNGAAGACCCAG 412  
117 LysValThrLeuAsnValHisSer...GluTyrGluProSerTrpGluGl 132  
|||||  
411 AAAGTCAACCTGAATGTCCACAGTGGTAGTCTCCCGCTGATTGGAGGC 362  
132 uGlnProMetProGluThrProLysTppPheHisLeuProTyrLeuPheG 149  
|||||  
361 TCAG.....ACTACCCAGAAATGTTTCACCTGGGAAGATTG...G 324  
149 InMetProAlaTyrAlaSerSer 156  
|||||  
323 AACAGCCATCTAAGTGAACAGT 301



Sequence	Strd. Orig	Zscore	EScore	Len	Documentation
gb_pr1:AF057557	+ 2047.00	1733.42	1.9e-88	1339	! AF057557 Homo sapiens anti-Pas
gb_btgl1:AC023534	+ 661.50	528.40	2.5e-21	221365	! AC023534 Homo sapiens clone
gb_btgt2:AC027719	- 284.00	220.68	0.0093	58628	! AC027719 Homo sapiens chrO3
gb_ro:MMU06431	+ 198.00	170.58	0.2098	3095	! U06431 Mus musculus B6/CBAJL
gb_ro:AF091137	+ 192.00	166.16	0.3699	2848	! AF091137 Trichosurus vulpecula
gb_ro:MMP1GIR03	+ 188.00	177.35	0.0860	417	! U83428 Mus musculus polymeric I
gb_ro:MMP1GIR3	+ 188.00	175.90	0.1061	505	! Y16525 Mus musculus pigR gene,
gb_ro:AE15253	+ 188.00	145.10	5.51	29392	! AE15253 Mouse pig receptor gen
gb_ro:AB001489	+ 188.00	145.10	5.51	29392	! AB001489 Mus musculus DNA for
gb_pat:EL15470	+ 187.50	165.39	0.4080	1911	! EL15470 cDNA encoding immunity
gb_ro:RNIGR	+ 185.50	159.64	0.8532	3269	! X15741 Rat mRNA for polymeric
gb_ro:AB0032195	+ 184.00	160.53	0.7613	2461	! AB0032195 Sus scrofa mRNA for H
gb_ro:BOVP1GIR	+ 183.00	156.96	1.20	3527	! U04797 Bos taurus polymeric im
gb_pr2:S43433503	+ 182.00	173.73	0.1400	345	! S43441 transmembrane secretary
gb_pat:AS2094	+ 182.00	160.66	0.7483	1936	! AS2094 Sequence 4 from Patent
gb_pat:AS2093	+ 182.00	160.30	0.7839	2031	! AS2093 Sequence 3 from Patent
gb_pat:AS2091	+ 182.00	159.02	0.9239	2405	! AS2091 Sequence 1 from Patent
gb_pr2:S62403	+ 182.00	157.61	1.11	2898	! S62403 transmembrane secretary
gb_pr1:HSPIR	+ 182.00	157.55	1.12	2919	! X73079 Homo sapiens encoding H
gb_ro:BP1GIR1	+ 182.00	149.30	1.38	3630	! X81371 B.taurus mRNA for bovin
gb_ro:AF035963	+ 171.00	145.95	3.39	2539	! AF035963 Rattus norvegicus kid
gb_btgl1:AC023534	- 167.00	112.12	378.14	221365	! AC023534 Homo sapiens clone
gb_ro:OCIGRECR	+ 166.50	143.09	7.13	3517	! X00412 Rabbit mRNA for poly-im
gb_bal:SC7H2	- 159.50	118.29	171.52	42655	! AL109732 Streptomyces coelic
gb_pr1:HUM1GPOLY	+ 157.00	137.60	14.41	2526	! M24559 Human poly-ig receptor
gb_ro:MMP1GIR5	+ 152.50	146.79	4.44	456	! Y16527 Mus musculus pigR gene,
gb_pr2:HS2238323	+ 152.50	141.07	9.23	969	! AJ738323 Homo sapiens mRNA for
gb_pr2:HS2010101	+ 152.50	141.07	9.24	970	! AJ010101 Homo sapiens mRNA for
gb_pr2:A22424864	+ 152.50	136.15	17.37	1857	! A7224864 Homo sapiens mRNA for
gb_pr2:HSU15177	+ 147.50	108.02	640.04	43599	! U15177 Human cosmid CRI-JC301
gb_btgt2:AL157395	+ 145.00	94.11	3.6e+03	207067	! AL157395 Homo sapiens chrom
gb_ro:MMP1GIR05	+ 143.00	138.38	13.03	481	! U83430 Mus musculus polymeric I
gb_btgt6:AC020151	- 142.50	102.95	1.2e+03	48836	! AC020151 Drosophila melanogas
gb_in2:AE003551	- 142.50	89.56	6.8e+03	259578	! AE003551 Drosophila melanoga
gb_bal:AB0032367	+ 142.00	100.37	1.7e+03	64957	! AB032367 Streptomyces avermit
gb_ba2:AE001986	- 141.50	110.03	494.93	17183	! AE001986 Deinococcus radiodur
gb_in1:LMP12743	+ 140.00	102.68	1.9e+03	43368	! AL031910 Leishmania major Fri
gb_bal:SCH35	+ 140.00	101.40	1.5e+03	45396	! AL078610 Streptomyces coelic
gb_btgt9:AC010208	+ 140.00	91.53	5.3e+03	167015	! AC010208 Homo sapiens clone
gb_btgt26:AC009813	- 140.00	90.07	6.4e+03	202598	! AC009813 Homo sapiens chrom

alignment\_block:

US-09-135-238B-2 x AF057557 ..

Align seg 1/1 to: AF057557 from: 1 to: 1339

```

1 MetAspArgTtpLeuTtpProLeuTyrPheLeuProValSerGlyAlaLe 17
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
20 ATGGAGCTTCCTGGCTTTGGCCACTTTACTTCTCCAGTATCGGGGGCCCT 69
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
17 uArgTlleLeuProGluValLysValGluGlyLeuGlyGlySerValT 34
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
70 GAGGATCCTCCAGAAAGTAAAGGTAGAGGGGAGCTGGCGGATCAGTTA 119
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
34 hrIleLysCysProLeuProGluMetHisValArgTlleTyrLeuCysArg 50
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
120 CCATCAAGTGGCCACTTCTGAAATGCATGTGAGGATATATCTGTGGCGG 169
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
51 GluMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsnPh 67
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
170 GAGATGGCTGGATCTGGAACATGTGTACCGTGTATCCACCACTT 219
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
67 eIleLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProAgl 84
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
220 CATCAAGCGCAATACAGAGGGCGAGTACTCTGAACATATACCCACGCA 269
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
84 ysAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerGly 100
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
270 AGAATCTGTCTTAGTGGAGGTAAACACAGCTGACAGAAAGTGACAGCGGA 319
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
101 ValTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLy 117
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
320 GTCTATGCTCGGAGCGGCATGACACAGACCGGGGAAGACCCAGAA 369
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
117 svalThrLeuAsnValHisSerGluTyrGluProSerTrrpGluGlnP 134
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
370 AGTCACCCCTGAATGTGCACAGTGAATACGAGGCATCATGGGAAGAGCAGC 419
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
134 roMetProGluThrProLysTrrpPheHisLeuProTyrLeuPheGlnMet 150
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
420 CAATGCTTGAGACTCCAAATGCTTTCATCTGCCCTATTTGTCAGATG 469
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
151 ProLafTyrAlaSerSerLysPheValThrArgValThrThrProAl 167
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
470 CCTGCATATGCCAGTTCTTCCAAATTCGTAACACAGAGTTACCCACACG 519
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
167 aGlnArgGlyLysValProProValHisHisSerSerProThrThrGlnI 184
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
520 TCAAGGGGGCAAGTCCCTCCAGTTCACCACTCTCTCCCCACCCACCAAA 569
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
184 leThrHisArgProArgValSerArgAlaSerSerValAlaGlyAspLys 200
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
570 TCACCCACCGCCCTCGAGTGTCCAGAGCATCTTCAGTAGCAGGTGACAG 619
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
201 ProArgThrPheLeuProSerThrThrAlaSerLysIleSerAlaLeuGl 217
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
620 CCCGCAACCTTCTCGCATCCACTACAGCCTCAGAAATCTCAGCTCTGGA 669
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
217 uGlyLeuLeuLysProGlnThrProSerTyrAsnHisThrArgLeuH 234
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
670 GGGGCTGCTCAAGCCCGAGAGCCGAGCTACAAACACACACACAGCGCTG 719
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
234 lsArgGlnArgAlaLeuAspTyrGlySerGlnSerGlyArgGluGlyGln 250
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
720 ACAGCAGAGAGCACTGGACTATGGCTACAGCTCTGGGAGGAAGGCCAA 769
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
251 GlyPheHisIleLeuIleProThrIleLeuGlyLeuPheLeuLeuAlaLe 267
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
770 GGAATTCACATCTCATCCGACCATCTCTGGCCCTTTTCTGCTGGCAT 819
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
267 uLeuGlyLeuValValLysArgAlaValGluArgArgLysAlaLeuSerA 284
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
820 TCTGGGCTGGTGGTGAAGAGGCCGTTGAAGAGGAGAAAGCCCTCTCCA 869

```

```

284 rgArgAlaArgArgLeuAlaValArgMetArgAlaLeuGluSerSerGln 300
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
870 GCGGGGCCCGCCGACTGGCGGTGAGGATGCGCGCCTGGAGAGTCCCGAG 919
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
301 ArgProArgGlySerProArgProArgSerGlnAsnAsnIleTyrSerAl 317
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
920 AGGCCCGCGGGTCCGCGGACCGCGCTCCCAAAACAACATCTACAGCGC 969
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
317 acYrProArgArgAlaArgGlyAlaAspAlaAlaGlyThrGlyGluAlap 334
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
970 CTGCGCGCGCGCTCGGTGGAGCGAGCGTGCAGGACACAGGGGAAGCCC 1019
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
334 roValProGlyProGlyAlaProLeuProAlaProLeuGlnValSer 350
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1020 CGGTTCGCGCGCGCGAGCGCGTTCGCGCGCGCGCGCGCTCGAGGTGCT 1069
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
351 GluSerProTrrpLeuHisAlapProSerLeuLysThrSerCysGluTyrVa 367
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1070 GAATCTCCCTGGCTCCATGCCCATCTCTGAAGACCGAGCTGTGAATACGT 1119
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
367 lSerLeuTyrHisGlnProAlaAlaMetMetGluAspSerAspSerAspA 384
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1120 GAGCCTCTACCAACCGCTGCGGCATGATGGAGGACAGTGTTCAGATG 1169
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
384 sPTyrIleAsnValProAla 390
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
1170 ACTACATCAATGTTCTCTGCC 1189
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
seq_name: gb_htg11:AC023534

seq documentation block:
LOCUS AC023534 221365 bp DNA HTG 02-MAR-2000
DEFINITION Homo sapiens clone RP11-462N18, *** SEQUENCING IN PROGRESS ***, 61
unordered pieces.
ACCESSION AC023534
VERSION AC023534.2 GI:7143452
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 221365)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome, clone RP11-462N18
Unpublished
2 (bases 1 to 221365)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,
Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castile,A.,
Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,
DeArelano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
Fenster,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Landers,T., Largoque,K., Lehoczy,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,E., Miranda,C., Mlenga,Y., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olivari,F.M.,
Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
Zody,M.
Direct Submission
Submitted (15-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 2, 2000 this sequence version replaced gi:6978210.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)

```

```
http://ftp.genome.washington.edu/RN/repeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center Project name: L5244
Center Clone name: 462_N_18
-----
* NONE: This is a 'working draft' sequence. It currently
* consists of 61 contigs. the true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1108: contig of 1108 bp in length
* 1109 1208: gap of 100 bp
* 1209 2282: contig of 1084 bp in length
* 2293 2392: gap of 100 bp
* 2393 3673: contig of 1281 bp in length
* 3674 3773: gap of 100 bp
* 3774 4947: contig of 1174 bp in length
* 4948 5047: gap of 100 bp
* 5048 6104: contig of 1057 bp in length
* 6105 6204: gap of 100 bp
* 6205 7259: contig of 1055 bp in length
* 7260 7359: gap of 100 bp
* 7360 8667: contig of 1308 bp in length
* 8668 8767: gap of 100 bp
* 8768 10566: contig of 1799 bp in length
* 10567 10666: gap of 100 bp
* 10667 12173: contig of 1507 bp in length
* 12174 12273: gap of 100 bp
* 12274 13643: contig of 1370 bp in length
* 13644 13743: gap of 100 bp
* 13744 15215: contig of 1472 bp in length
* 15216 15315: gap of 100 bp
* 15316 16609: contig of 1294 bp in length
* 16610 16709: gap of 100 bp
* 16710 18085: contig of 1376 bp in length
* 18086 18185: gap of 100 bp
* 18186 19668: contig of 1483 bp in length
* 19669 19768: gap of 100 bp
* 19769 22083: contig of 2315 bp in length
* 22084 22183: gap of 100 bp
* 22184 24052: contig of 1869 bp in length
* 24053 24152: gap of 100 bp
* 24153 25902: contig of 1750 bp in length
* 25903 26002: gap of 100 bp
* 26003 27639: contig of 1637 bp in length
* 27640 27739: gap of 100 bp
* 27740 29377: contig of 1638 bp in length
* 29378 29477: gap of 100 bp
* 29478 30910: contig of 1433 bp in length
* 30911 31010: gap of 100 bp
* 31011 31225: contig of 215 bp in length
* 31226 31325: gap of 100 bp
* 31326 32890: contig of 1565 bp in length
* 32891 32990: gap of 100 bp
* 32991 34812: contig of 1822 bp in length
* 34813 34912: gap of 100 bp
* 34913 36817: contig of 1905 bp in length
* 36818 36917: gap of 100 bp
* 36918 38575: contig of 1658 bp in length
* 38576 38675: gap of 100 bp
* 38676 41309: contig of 2634 bp in length
* 41310 41409: gap of 100 bp
* 41410 43344: contig of 1935 bp in length
* 43345 43444: gap of 100 bp
* 43445 45853: contig of 2409 bp in length
*
* 45854 45953: gap of 100 bp
* 45954 48434: contig of 2481 bp in length
* 48435 48534: gap of 100 bp
* 48535 51026: contig of 2492 bp in length
* 51027 51126: gap of 100 bp
* 51127 54369: contig of 3243 bp in length
* 54370 54469: gap of 100 bp
* 54470 56661: contig of 2192 bp in length
* 56662 56761: gap of 100 bp
* 56762 59097: contig of 2336 bp in length
* 59098 59197: gap of 100 bp
* 59198 61551: contig of 2354 bp in length
* 61552 61651: gap of 100 bp
* 61652 64257: contig of 2606 bp in length
* 64258 64357: gap of 100 bp
* 64358 67511: contig of 3154 bp in length
* 67512 67611: gap of 100 bp
* 67612 70329: contig of 2718 bp in length
* 70330 70429: gap of 100 bp
* 70430 73572: contig of 3143 bp in length
* 73573 73672: gap of 100 bp
* 73673 77203: contig of 3531 bp in length
* 77204 77303: gap of 100 bp
* 77304 80599: contig of 3296 bp in length
* 80600 80699: gap of 100 bp
* 80700 84046: contig of 3347 bp in length
* 84047 84146: gap of 100 bp
* 84147 88357: contig of 4211 bp in length
* 88358 88457: gap of 100 bp
* 88458 92454: contig of 3997 bp in length
* 92455 92554: gap of 100 bp
* 92555 96892: contig of 4338 bp in length
* 96893 96992: gap of 100 bp
* 96993 101334: contig of 4342 bp in length
* 101335 101434: gap of 100 bp
* 101435 105438: contig of 4004 bp in length
* 105439 105539: gap of 100 bp
* 105539 110413: contig of 4875 bp in length
* 110414 110513: gap of 100 bp
* 110514 115994: contig of 5481 bp in length
* 115995 116094: gap of 100 bp
* 116095 120447: contig of 4353 bp in length
* 120448 120547: gap of 100 bp
* 120548 124483: contig of 3936 bp in length
* 124484 124583: gap of 100 bp
* 124584 130249: contig of 5666 bp in length
* 130250 130349: gap of 100 bp
* 130350 135462: contig of 5113 bp in length
* 135463 135562: gap of 100 bp
* 135563 139847: contig of 4285 bp in length
* 139848 139947: gap of 100 bp
* 139948 146731: contig of 6784 bp in length
* 146732 146831: gap of 100 bp
* 146832 151469: contig of 4638 bp in length
* 151470 151569: gap of 100 bp
* 151570 159901: contig of 8332 bp in length
* 159902 160001: gap of 100 bp
* 160002 168308: contig of 8307 bp in length
* 168309 168408: gap of 100 bp
* 168409 176195: contig of 7787 bp in length
* 176196 176295: gap of 100 bp
* 176296 189303: contig of 13008 bp in length
* 189304 189403: gap of 100 bp
* 189404 202135: contig of 12732 bp in length
* 202136 202235: gap of 100 bp
* 202236 221365: contig of 19130 bp in length.
*
* FEATURES
*   Location/Qualifiers
*     1..221365
*       /organism="Homo sapiens"
*       /db_xref="taxon:9606"
*       /clone="RP11-462N18"
*       /clone_lib="RPC1-11 Human Male BAC"
*     misc_feature
*       1..1108
```

```

misc_feature      /note="assembly_fragment"
1209..2292
misc_feature      /note="assembly_fragment"
2393..3673
misc_feature      /note="assembly_fragment"
3774..4947
misc_feature      /note="assembly_fragment"
5048..6104
misc_feature      /note="assembly_fragment"
6205..7259
misc_feature      /note="assembly_fragment"
7360..8667

alignment_scores:
  Quality: 661.50      Length: 508
  Ratio: 3.410        Gaps: 8
Percent Similarity: 38.189  Percent Identity: 34.252

alignment_block:
US-09-135-238B-2 x AC023534 ..
Align seg 1/1 to: AC023534 from: 1 to: 221365

2 AspArgTyrLeuTrp..... 6
33094 GACGGCTGGAGCTGGGAAGTAGATGGGCTGGGATATAGTTAGAGATGGG 33143
|||||
7 ProLeuTyrPheLeuPro..... ValSerGlyAlaLeuArgI 19
33144 GCTCCTCTGCAATTAATCTCTCTCCCTGCGAGTATCGGGGGCCCTGAGGA 33193
|||||
19 LeuProGluValIysValGluLeuLeuGlyGlySerValThrIle 35
33194 TCCTCCAGAAAGTAGAGGGGAGCTGGCGGATCAGTTACCATC 33243
|||||
36 LysCysProGluMetHisValArgIleTyrLeuCysArgGluMe 52
33244 AAGTGCCCACTTCCTGAATGCAATGAGGATATATCTGTGCGGGAGAT 33293
|||||
52 tAlaGlySerGlyThrCysGlyThrValValSerThrThrAspPheIleL 69
33294 GGCTGGATCTGGAACATGTGGTACCGTGTATCCACCACCACTTCATCA 33343
|||||
69 yAlaGluTyrIysGlyArgValThrLeuLysGlnTyrProArgLysAsn 85
33344 AGCAGAATACAGGGCGGAGTTACTCTGAAGCAATACCCACGCAAGAT 33393
|||||
86 LeuPheLeuValGluValThrGlnLeuThrGluSerAspSerGlyValTy 102
33394 CTGTTCTAGTGGAGTAAACACAGCTGACAGAAAGTCACAGCGGAGTCTA 33443
|||||
102 rAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLysValT 119
33444 TGCTCGGAGGGGCGCATGAACACAGACCGGGGAAGACCCAGAAAGTCA 33493
|||||
119 hrLeuAsnValHis..... 123
33494 CCCTGAATGTGCACAGTGGTAGGTTCTCTCCGCTGATTGGAGGCTCAGACT 33543
|||||
123 ..... 123
33544 ACCCAGAAATGTTTCATCTGGGAAGATAGGAACAGACATCTAACTGTGA 33593
|||||
123 ..... 123
33594 CAGAGTCTCTAGACTCTTACACTGTGGCCATATAAGCCACATAATGATTGA 33643
|||||
123 ..... 123
33644 GATCCATATGAACCTCATAGTCAATTACCTCCATGTGTCCACACAATTTT 33693
|||||
123 ..... 123

33694 CTACCCACAATGATATGAAGAACAAGATAGGTGCTCTCTTTGTTATGTTT 33743
123 ..... 123
33744 CATCTTACCAGGCATGAGGGGATAGAGGCTTGAAGATTGAGAGAGGC 33793
123 ..... 123
33794 TCTGAGATGGAGAAATTGGAAGCAAAAGGAATAACTATCAACGTAAT 33843
123 ..... 123
33844 TGTCCCATGAGAAATAAGAGAAAGACACTAATTTATATCCAGCCACATTTAT 33893
123 ..... 123
33894 AGGCCAATCTTATTACTCTGTGACCTGATCTTTATTATTCTCTGGAACCC 33943
123 ..... 123
33944 GCCTTTCTCTCTCTCCACTGAGTTCCTCCACACATATGCACAGTAATT 33993
123 ..... 123
33994 CAACAGCCTAGAAATCCCAATGACCTCTGCTGTTTTCAGGGAATGTTGA 34043
123 ..... 123
34044 CTGACTAGGGGAACCAAGTATTTCTCCAGGAAGCCCCCAACCCAGCAGTC 34093
123 ..... 123
34094 CTCTCTGCTCTTTTCTGAGTGGACCTGCTCCCTCATCCCTCACTGTCATCC 34143
123 ..... 123
34144 ACTTTGGCTGAATTAGCTTCAGTTTACACCCAGCCCTCAGTCTCCCCCTA 34193
124 ..... Ser..GluTyrGluProSerTr 130
34194 GAGACCACCAACATATTCTCTCTCTTCCATAGATAAGGACCATCATG 34243
|||||
130 pGluGlnProMetProGluThrProLysTrpPheHisLeuProTyrIL 147
34244 GGAAGACAGCCATGCTGAGACTCCAAATGGTTTCTATCTGCCCTATT 34293
|||||
147 euPheGlnMetProAlaTyrAlaSerSerLysPheValThrArg... 162
34294 TGTTCAGATGCTGCATATGCCAGTTCTTCCAAATTCGTAACCAAGGT 34343
|||||
163 ValThrThrProAlaGlnArgGlyLysValPro..... 173
34344 CAGTTCCCATNAGCAAGGTTAGGAGAGTGGGTGAGATGCTGTGGCAATG 34393
|||||
174 ..... ProValHisHisSerSerProThrT 182
34394 TGAATAATGGAGAGGGGCTATGTTTCAGTTTCAT..... 34427
182 hrGlnIleThrHisArgProArgValSerArgAlaSerSerValAlaGly 198
34428 ..AGATGGCTGGCAGAGCTGAGTCTTAGGTGTGTATTCCTGCC 34475
|||||
199 AspLysProArgThrPheLeuProSerThrThrAlaSerLysIleSerAl 215
34476 TTTAGAGGGCACATACCTCCCTGAGATCTGGGGTCCATATCTCTTCAAG 34525
|||||
215 aLeuGluGlyLeuLeuLysProGlnThrProSerTyrAsnHisHis.... 230
34526 CCTGTCGACAGCTCAGGTCCCTGGTCTCTAGTGCAGAACTTACATATTA 34575
|||||
231 ..ThrArgLeuHisArgGlnArg 237
34576 CCACAGGTGGCACCACCAAGCAA 34598

```



```

seq_name: gb_htg22:AC027719
seq_documentation_block:
LOCUS      AC027719      58628 bp      DNA      HTG      01-APR-2000
DEFINITION Homo sapiens chromosome 8 clone RP11-258P9 map 8, LOW-PASS SEQUENCE
SAMPLING.
ACCESSION  AC027719
VERSION    AC027719.1 GI:7382531
KEYWORDS  HTG; HTGS_PHASE0.
SOURCE    human.

ORGANISM  Homo sapiens
           Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 58628)
AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE     Homo sapiens chromosome 8, clone RP11-258P9
JOURNAL   Unpublished
REFERENCE  2 (bases 1 to 58628)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
           Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
           Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
           Campopiano,A., Castile,A., Choepel,Y., Colangelo,M., Collins,S.,
           Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
           Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
           Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
           Grand-Pierre,N., Grant,G., Hagos,B., Headford,A., Horton,L.,
           Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
           Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczyk,J.,
           Levine,R., Lieu,C., Liu,C., Liu,C., Locke,K., Macdonald,P., Marquis,N.,
           McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
           Melarim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
           Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
           O'Neill,D., Ollivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
           Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
           Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
           Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
           Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
           Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
           Young,G., Zalnoun,J., Zimmer,A. and Zody,M.
           Direct Submission
           Submitted (01-APR-2000) Whitehead Institute/MIT Center for Genome
           Research, 320 Charles Street, Cambridge, MA 02141, USA
           All repeats were identified using RepeatMasker:
           Smit, A.F.A. & Green, P. (1996-1997)
           http://ftp.genome.washington.edu/RM/RepeatMasker.html
           ----- Genome Center
           Center: Whitehead Institute/ MIT Center for Genome Research
           Center code: WIBR
           Web site: http://www-seq.wi.mit.edu
           Contact: sequence.submissions@genome.wi.mit.edu
           ----- Project Information
           Center project name: L7449
           Center clone name: 258_P_9
           -----
           * NOTE: This record contains 74 individual
           * sequencing reads that have not been assembled into
           * contigs. Runs of N are used to separate the reads
           * and the order in which they appear is completely
           * arbitrary. Low-pass sequence sampling is useful for
           * identifying clones that may be gene-rich and allows
           * overlap relationships among clones to be deduced.
           * However, it should not be assumed that this clone
           * will be sequenced to completion. In the event that
           * the record is updated, the accession number will
           * be preserved.
           *
           * 1 671: contig of 671 bp in length
           * 672 771: gap of 100 bp
           * 772 1469: contig of 698 bp in length
           * 1470 1569: gap of 100 bp
           * 1570 2261: contig of 692 bp in length
           * 2262 2361: gap of 100 bp
           * 2362 3051: contig of 690 bp in length
           *
           * 3052 3151: gap of 100 bp
           * 3152 3816: contig of 665 bp in length
           * 3817 3916: gap of 100 bp
           * 3917 4597: contig of 681 bp in length
           * 4598 4697: gap of 100 bp
           * 4698 5395: contig of 698 bp in length
           * 5396 5495: gap of 100 bp
           * 5496 6185: contig of 690 bp in length
           * 6186 6285: gap of 100 bp
           * 6286 6969: contig of 684 bp in length
           * 6970 7069: gap of 100 bp
           * 7070 7780: contig of 711 bp in length
           * 7781 8590: contig of 710 bp in length
           * 8591 8690: gap of 100 bp
           * 8691 9372: contig of 682 bp in length
           * 9373 9472: gap of 100 bp
           * 9473 10162: contig of 690 bp in length
           * 10163 10262: gap of 100 bp
           * 10263 10944: contig of 682 bp in length
           * 10945 11044: gap of 100 bp
           * 11045 11747: contig of 703 bp in length
           * 11748 11847: gap of 100 bp
           * 11848 12547: contig of 700 bp in length
           * 12548 12647: gap of 100 bp
           * 12648 13336: contig of 689 bp in length
           * 13337 13436: gap of 100 bp
           * 13437 14103: contig of 667 bp in length
           * 14104 14203: gap of 100 bp
           * 14204 14872: contig of 669 bp in length
           * 14873 14972: gap of 100 bp
           * 14973 15682: contig of 710 bp in length
           * 15683 15782: gap of 100 bp
           * 15783 16491: contig of 709 bp in length
           * 16492 16591: gap of 100 bp
           * 16592 17290: contig of 699 bp in length
           * 17291 17390: gap of 100 bp
           * 17391 18088: contig of 698 bp in length
           * 18089 18188: gap of 100 bp
           * 18189 18883: contig of 695 bp in length
           * 18884 18983: gap of 100 bp
           * 18984 19671: contig of 688 bp in length
           * 19672 19771: gap of 100 bp
           * 19772 20413: contig of 642 bp in length
           * 20414 20513: gap of 100 bp
           * 20514 21196: contig of 683 bp in length
           * 21197 21296: gap of 100 bp
           * 21297 21992: contig of 696 bp in length
           * 21993 22092: gap of 100 bp
           * 22093 22804: contig of 712 bp in length
           * 22805 22904: gap of 100 bp
           * 22905 23604: contig of 700 bp in length
           * 23605 23704: gap of 100 bp
           * 23705 24402: contig of 698 bp in length
           * 24403 24502: gap of 100 bp
           * 24503 25191: contig of 689 bp in length
           * 25192 25291: gap of 100 bp
           * 25292 25974: contig of 683 bp in length
           * 25975 26074: gap of 100 bp
           * 26075 26765: contig of 695 bp in length
           * 26770 26869: gap of 100 bp
           * 26870 27572: contig of 703 bp in length
           * 27573 27672: gap of 100 bp
           * 27673 28373: contig of 701 bp in length
           * 28374 28473: gap of 100 bp
           * 28474 29165: contig of 692 bp in length
           * 29166 29265: gap of 100 bp
           * 29266 29956: contig of 691 bp in length
           * 29957 30056: gap of 100 bp
           * 30057 30741: contig of 685 bp in length
           * 30742 30841: gap of 100 bp
           * 30842 31519: contig of 678 bp in length
           * 31520 31619: gap of 100 bp

```

31620 32323: contig of 704 bp in length  
32324 32423: gap of 100 bp  
32424 33110: contig of 687 bp in length  
33111 33210: gap of 100 bp  
33211 33902: contig of 692 bp in length  
33903 34002: gap of 100 bp  
34003 34708: contig of 706 bp in length  
34709 34808: gap of 100 bp  
34809 35500: contig of 692 bp in length  
35501 35600: gap of 100 bp  
35601 36300: contig of 700 bp in length  
36301 36400: gap of 100 bp  
36401 37097: contig of 697 bp in length  
37098 37197: gap of 100 bp  
37198 37900: contig of 703 bp in length  
37901 38000: gap of 100 bp  
38001 38695: contig of 695 bp in length  
38696 38795: gap of 100 bp  
38796 39487: contig of 692 bp in length  
39488 39587: gap of 100 bp  
39588 40291: contig of 704 bp in length  
40292 40391: gap of 100 bp  
40392 41101: contig of 710 bp in length  
41102 41201: gap of 100 bp  
41202 41886: contig of 685 bp in length  
41887 41986: gap of 100 bp  
41987 42686: contig of 700 bp in length  
42687 42786: gap of 100 bp  
42787 43484: contig of 698 bp in length  
43485 43584: gap of 100 bp  
43585 44276: contig of 692 bp in length  
44277 44376: gap of 100 bp  
44377 45078: contig of 702 bp in length  
45079 45178: gap of 100 bp  
45179 45869: contig of 691 bp in length  
45870 45969: gap of 100 bp  
45970 46664: contig of 695 bp in length  
46665 46764: gap of 100 bp  
46765 47456: contig of 692 bp in length  
47457 47556: gap of 100 bp  
47557 48249: contig of 693 bp in length  
48250 48349: gap of 100 bp  
48350 49042: contig of 693 bp in length  
49043 49142: gap of 100 bp  
49143 49843: contig of 701 bp in length  
49844 49943: gap of 100 bp  
49944 50640: contig of 697 bp in length  
50641 50740: gap of 100 bp  
50741 51438: contig of 698 bp in length  
51439 51538: gap of 100 bp  
51539 52231: contig of 693 bp in length  
52232 52331: gap of 100 bp  
52332 53020: contig of 689 bp in length  
53021 53120: gap of 100 bp  
53121 53857: contig of 737 bp in length  
53858 53957: gap of 100 bp  
53958 54652: contig of 695 bp in length  
54653 54752: gap of 100 bp

alignment\_scores:

Quality: 284.00 Length: 80  
Ratio: 4.733 Gaps: 2  
Percent Similarity: 75.000 Percent Identity: 73.750

alignment\_block:

US-09-135-238b-2 x AC027719/rev ..

Align seg 1/1 to reverse of: AC027719 from: 1 to: 58628

2 AspArgTrpLeuPrp..... 6

40676 GACGCGTGGAGCTGGGAAGTAGATGGGCTGGGATATAGTTAGATGGG 40627

7 ....ProLeuTyrPheLeuPro.....ValSerGlyAlaLeuArgI 19  
|||||:|:| ||| |||||  
40626 GCTCCTCTGCATTAACTCTCCTCCTCGAGTATCATGGGGCCCTGAGGA 40577  
19 leLeuProGluValLysValGluGlyGluLeuGlyGlySerValThrIle 35  
|||||  
40576 TCTTCCCAAGAGTAAGGTAGAGGGGAGCTGGGGGATCAGTTACCATC 40527  
36 LysCysProLeuProGluMetHisValArgIleTyrLeuCysArgGluMe 52  
|||||  
40526 AAGTGGCCACTCTCTGAAATGCATGTGAGGATATATCTGTGCCGGGAGAT 40477  
52 talacGlySerGlyThrCysGlyThrValValSerThrThr 65  
|||||  
40476 GCGTGGATCTGGAAATGTGGTACCGTGGTATCCACCACC 40437  
seq\_name: gb\_ro:MMU06431  
seq\_documentation\_block: 3095 bp mRNA ROD 26-MAY-1995  
LOCUS MMU06431  
DEFINITION Mus musculus B6/CBAF1J polymeric immunoglobulin receptor mRNA,  
complete cds.  
ACCESSION U06431  
VERSION U06431.1 GI:458421  
KEYWORDS .  
SOURCE mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 3095)  
AUTHORS Piskurich,J.F., Blanchard,M.H., Youngman,K.R., France,J.A. and  
Kaetzel,C.S.  
TITLE Molecular cloning of the mouse polymeric Ig receptor. Functional  
regions of the molecule are conserved among five mammalian species  
J. Immunol. 154 (4), 1735-1747 (1995)  
MEDLINE 95138517  
REFERENCE 2 (sites)  
AUTHORS Piskurich,J.F., Hsieh,M., Youngman,K.R. and Kaetzel,C.S.  
TITLE Molecular cloning of mouse polymeric Ig receptor  
J. Immunol. 150, 38A-38A (1993)  
REFERENCE 3 (bases 1 to 3095)  
AUTHORS Piskurich,J.F.  
TITLE Molecular Cloning and Regulation of the Polymeric Immunoglobulin  
Receptor  
Thesis (1994) Pathology, Case Western Reserve University  
REFERENCE 4 (bases 1 to 3095)  
AUTHORS Piskurich,J.F.  
TITLE Direct Submission  
Submitted (08-FEB-1994) Janet F. Piskurich, Case Western Reserve  
University, Pathology, Biomedical Research Building, Cleveland, OH  
44060, USA  
FEATURES  
Location/Qualifiers  
1..3095  
/organism="Mus musculus"  
/strain="B6/CBAF1J (C57 Black 6 X CBA)"  
/db\_xref="taxon:10090"  
/clone="G20"  
/clone\_lib="catalog #935302, Stratagene Cloning Systems,  
La Jolla, CA"  
/sex="female"  
/tissue\_type="liver"  
5'UTR <1..84  
sig\_peptide 85..138  
CDS 85..2400  
/codon\_start=1  
/function="binds and transports polymeric immunoglobulin"  
/product="polymeric immunoglobulin receptor"  
/protein\_id="AAA67440.1"  
/db\_xref="GI:458422"  
/translation="MRLLFTLLVTVFGVSTKSPFGQEVSSIEGDSVITCYTPD  
TSVNRHTKRYWCROGAGCTTLLISNGYLSKEYSGRANLINFNNFTVINLEOLTQ  
DDTGSYKGLGTSNRGLSDVLSYQVPELSDTHVTKDIGRNVITIECFPRENVF  
SKKSCKKTNQSCELVIDSTKVNPSYIGRAKLFMKGTDLTFTVYVNIISHLTHNDAGLY





/gene="pigR"  
/label=ex3  
/usedin=y16524:pig\_cds  
/usedin=y16524:pig\_mrna  
BASE COUNT 130 a 136 c 127 g 112 t  
ORIGIN

alignment\_scores:  
Quality: 188.00 Length: 97  
Ratio: 2.648 Gaps: 3  
Percent Similarity: 73.196 Percent Identity: 40.206

alignment\_block:

US-09-135-238B-2 x MPMIGR3 ..

Align seg 1/1 to: MPMIGR3 from: 1 to: 505

30 GlyGlySerValThrIleLeuLysCysProLeuProGlu.....Me 42  
||| |||||:|||| ||| |||||  
130 GCGGACTCTGTTCCATCAGCTGCTACTACCCAGACACCTCTGTCAACCG 179  
42 ThisValArgIleTyLeuCysArgGluMetAlaGlySerGlyThrCysG 59  
|||:||||| ||| |||||: ||| ||||| |||  
180 GCACACCCGGAATACTGGTCCGCAAGAGGCC....AGCGGCATGTGCA 226  
59 lyThrValValSerThrThrAsnPhelLeuLysAlaGluTyLysGlyArg 75  
|||:|||||:|||||:|||||: |||||: |||||  
227 CAACGCTCATCTCTCAATGGCTACCTCTCCAGGAGTATTAGGCAGA 276  
76 ValThrLeuLysGlnTyProArgLysAsnLeuPheLeuValGluValTh 92  
:||||| |||:|||||: ||| |||||: |||||  
277 GCCAACCTCATCACTCCAGAGAACACACATTTGTGATTAACTTGA 326  
92 rGlnLeuThrGluSerAspSerGlyValTyAlaCysGlyAlaGlyMeta 109  
:||||| |||||: ||| |||||: |||||  
327 GCAGCTCACCAGGAGCTGCTTCGATGTCACGCTGGAGGTC 414  
109 snThrAspArgGlyLysThrGlnLysValThrLeuAsnVal 122  
:||||| |||||: ||| |||||: |||||  
374 CCAGTAACCGAGGCTGCTTCGATGTCACGCTGGAGGTC 414

seq\_name: gb\_pat:E15253

seq\_documentation\_block:

LOCUS E15253 29392 bp DNA PAT 28-JUL-1999  
DEFINITION Mouse pig receptor gene.  
ACCESSION E15253  
VERSION E15253.1 GI:5709936  
KEYWORDS JP 1998057066-A/1.  
SOURCE Mus sp.  
ORGANISM Mus sp.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

Sato,T. and Kushi,A..

MURINE PIG RECEPTOR PROTEIN GENE

PATENT: JP 1998057066-A 03-MAR-1998;  
YAKULT HONSHA CO LTD

OS Mus sp. (mouse)

PN JP 1998057066-A/1

PD 03-MAR-1998

PF 19-AUG-1996 JP 1996217154

PI SATO TAKASHI, KUSHIRO AKIRA

PC C12N15/09,C07H21/04,C07K14/705;

CC strandedness: Double;

CC topology: Linear;

EH key Location/Qualifiers

FT

FT source

1. .29392

/organism="Mus sp."

/clone\_lib="genomic DNA from female 129SVJ FT

mouse liver in

FEATURES  
source  
Location/Qualifiers  
1. .29392  
/organism="Mus sp."  
/db\_xref="taxon:10095"  
BASE COUNT 8318 a 6747 c 6514 g 7813 t  
ORIGIN

alignment\_scores:  
Quality: 188.00 Length: 97  
Ratio: 2.648 Gaps: 3  
Percent Similarity: 73.196 Percent Identity: 40.206

alignment\_block:

US-09-135-238B-2 x E15253 ..

Align seg 1/1 to: E15253 from: 1 to: 29392

30 GlyGlySerValThrIleLeuLysCysProLeuProGlu.....Me 42  
||| |||||:|||| ||| |||||  
18245 GCGGACTCTGTTCCATCAGCTGCTACTACCCAGACACCTCTGTCAACCG 18294  
42 ThisValArgIleTyLeuCysArgGluMetAlaGlySerGlyThrCysG 59  
|||:||||| ||| |||||: ||| ||||| |||  
18295 GCACACCCGGAATACTGGTCCGCAAGAGGCC....AGCGGCATGTGCA 18341  
59 lyThrValValSerThrThrAsnPhelLeuLysAlaGluTyLysGlyArg 75  
|||:|||||:|||||:|||||: |||||: |||||  
18342 CAACGCTCATCTCTCAATGGCTACCTCTCCAGGAGTATTAGGCAGA 18391  
76 ValThrLeuLysGlnTyProArgLysAsnLeuPheLeuValGluValTh 92  
:||||| |||:|||||: ||| |||||: |||||  
18392 GCCAACCTCATCACTCCAGAGAACACACATTTGTGATTAACTTGA 18441  
92 rGlnLeuThrGluSerAspSerGlyValTyAlaCysGlyAlaGlyMeta 109  
|||:|||||:|||||: ||| |||||: |||||  
18442 GCAGCTCACCAGGAGCAGCTGGGAGCTACAAGTGTGCGCTGGGT...A 18488  
109 snThrAspArgGlyLysThrGlnLysValThrLeuAsnVal 122  
:||||| |||||: ||| |||||: |||||  
18489 CCAGTAACCGAGGCTGCTTCGATGTCACGCTGGAGGTC 18529

seq\_name: gb\_to:AB001489

seq\_documentation\_block:

LOCUS AB001489 29392 bp DNA ROD 14-APR-2000  
DEFINITION Mus musculus DNA for polyimmunoglobulin receptor, complete cds.  
ACCESSION AB001489  
VERSION AB001489.1 GI:2804245  
KEYWORDS polyimmunoglobulin receptor.  
SOURCE Mus musculus (Strain:129SVJ) female liver DNA.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (sites)

Kushiro,A. and Sato,T.

Polymeric immunoglobulin receptor gene of mouse: sequence,

structure and chromosomal location

Gene 204 (1-2), 277-282 (1997)

98094283

REFERENCE 2 (bases 1 to 29392)

Sato,T.

Direct Submission

Submitted (26-FEB-1997) to the DDBJ/EMBL/GenBank databases. Takashi

Sato, Yakult Central Institute for Microbiological Research,

Culture Collection Center; 1796 Yaho, Kunitachi-shi, Tokyo 186,

Japan (Tel:425-77-8960, Fax:425-77-3020)

Location/Qualifiers

1. .29392

/organism="Mus musculus"

/strain="129SVJ"

/db\_xref="taxon:10090"

```
/chromosome="1"
/map="IE3"
/sex="female"
/tissue_type="liver"
2470. .2474
2580. .2585
2878. .2882
3176. .3180
3251. .3255
3350. .3355
3389. .3501
/feature="based on the putative transcription start"
3389. .3501
/number=1
3502. .11067
/number=1
11124. .27591
/feature="pigr"
23908. .24240,25133. .25325,25686. .25813,26183. .26314,
26426. .26484,27496. .27591)
/feature="pigr"
/feature="mediates secretion of IgA and IgM"
/codon_start=1
/product="polyimmunoglobulin receptor"
/protein_id="BAA24431.1"
/db_xref="GI:2804246"
/translation="MRLILFTLLVTFVSGVSKSPIFGPOVSSIEGDSVITCYIPD
TSVNRHRYWCROGAGMCTLLISSNGYLSKEYSGRANLFPENNIFVINIQLPQ
DDTGYKCLGTSNRGLSFDVLSVQVPELPSDTHVTYTDIGRNVITECFKRENA
SKSLACKLTNQSCELVIDSTEKNVPSYIGRAKLFMGKTDLTVPYINISHLTHDAGLY
ICQAGPEPSADKKNDVLQVLAPELLYKDLRSLVTECDLGRVANAAYLCRMNKE
TCVDILNTLGRKDPDFEGRILITPKDDNGRFSVLITGLRKEDAGHVGCAHSSGLPOE
GNPLOTWQLPVNEESTIPNRRSVYKGTGGSVAACYPNPKESSILKYWCWEGDNG
HCPVLVGTQVQVQVEIEGRALPDQFENGITVIVLNQLTIDAGFTWLTNGDSNRRT
TIELQVAEATREPNEVLTPQNAVATLGETFTVSHYPCKFYSOEKRYCKWKNKCHIL
PSHDEGARQSSVSCDQSSQLVSMTLNPVSKDEGNTWCVKQKQGTGETTAYIAVEE
RTGSHVNPTDANARAKVALEEVVDSSISEKENKALPNPFPANEREIQNVGDQAO
ENRAGSDAGSADQSGRSSSKVLFTVLPLGLVLAAGIAVAVRVRHKKVNDKMSIS
SVRTDLSMADFKNSDLGNDNMGASPD/QQTVIEGRDEIVTTTCTAEPESKKAKR
SSKREADMAYSAPLQSSSTIAQVHDGPQEA"
11124. .11166
/feature="pigr"
/number=2
11167. .18191
/feature="pigr"
/number=2
18192. .18536
/feature="pigr"
/number=3
18537. .21129
/feature="pigr"
/number=3
21130. .21783
/feature="pigr"
/number=4
21784. .23021
/feature="pigr"
/number=4
23022. .23354
/feature="pigr"
/number=5
23355. .23907
/feature="pigr"
/number=5
23908. .24240
/feature="pigr"
/number=6
24241. .25132
/feature="pigr"
/number=6
25133. .25325
/feature="pigr"
```

```
intron
25326. .25685
/feature="pigr"
/number=7
25686. .25813
/feature="pigr"
/number=8
25814. .26182
/feature="pigr"
/number=8
26183. .26314
/feature="pigr"
/number=9
26315. .26425
/feature="pigr"
/number=9
26426. .26484
/feature="pigr"
/number=10
26485. .27495
/feature="pigr"
/number=10
27496. .29008
/number=11
27592. .29008
/feature="based on the polyadenylation site"
polya_signal
28983. .28988
polya_site
29008
BASE COUNT 8318 a 6747 c 6514 g 7813 t
ORIGIN

alignment_scores:
Quality: 188.00 Length: 97
Ratio: 2.648 Gaps: 3
Percent Similarity: 73.196 Percent Identity: 40.206

alignment_block:
US-09-135-238B-2 x AB001489 ..
Align seg 1/1 to: AB001489 from: 1 to: 29392

30 GlyCysSerValThrIleLysCysProLeuProGlu.....Me 42
||| |||||:|||| ||| |||||:
18245 GCGGACTCTGTTTCATCGTCTACTACCCAGACACCTCTGTCAACCG 18294

42 tHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 59
|||:|||| ||| |||||: ||| ||||| |||
18295 GCACACCCGGAATACTGTGCGCAGACAGAGGCC...AGCGCATGTGCA 18341

59 lyThrValValSerThrThrAsnPhelLeLysAlaGluTyrLysGlyArg 75
|||||:|||||:|||||:|||||: |||||:|||||
18342 CAACGCTCATCTCTCAATGGCTACTCTCCAGAGGATGATTACGACAGA 18391

76 ValThrLeuLysGlnTyrProArgLysAsnLeuPhelLeuValGluValTh 92
:|||||: |||:|||||: |||:|||||: |||:|||||:
18392 GCCAACTCATCACTCCAGAGAACACACATTTGTGATTACATTGA 18441

92 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 109
|||||:|||||:|||||: ||| ||||| |||
18442 GCAGCTCACCCAGGACGACTGGGAGCTACAAAGTGTGGCTGGGT...A 18488

109 snThrAspArgGlyLysThrGlnLysValThrLeuAsnVal 122
:|||||:|||||: |||: |||||:|||||:
18489 CCAGTACCGAGCGCTGCTCTTCGATGTCAGCTGGAGGTC 18529

seq_name: gb_pat:E15470

seq_documentation_block:
LOCUS E15470 1911 bp DNA PAT 28-JUL-1999
DEFINITION CDNA encoding immunity related factor.
ACCESSION E15470
VERSION E15470.1 GI:5710153
```

```

KEYWORDS  JP 1998072495-A/1.
SOURCE     Homo sapiens.
ORGANISM   Homo sapiens.
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS    Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE      1 (bases 1 to 1911)
JOURNAL    Miyabayashi,T. and Sakano,S. .
IMMUNITY-RELATED FACTOR
PATENT:    JP 1998072495-A 17-MAR-1998;
ASNTI CHEM IND CO LTD
OS         Homo sapiens (human)
PN         JP 1998072495-A/1
PD         17-MAR-1998
PR         11-JUN-1997 JP 1997153218
PR         13-JUN-1996 JP 96P 152362
PI         MIYABAYASHI TOMOYUKI, SAKANO SEIJI
PC         C07K14/47,C07H21/04,C07K16/18,C12N5/10,C12N15/09, PC
C12P21/02//A61K38/00,
PC         C12P21/08,(C12N5/10,C12R1:91),(C12P21/02,C12R1:91),(C12P21/08,
PC         C12R1:91);
CC         strandedness: Double;
CC         topology: Linear;
CC         hypothetical: No;
CC         anti-sense: No;
FH         Key
FH         Location/Qualifiers
FT         1..1911
FT         /organism='Homo sapiens'
FT         /tissue_type='Lymph node'
FT         5'UTR
FT         1..97
FT         sig_peptide
FT         98..145
FT         /product='signal peptide of immunity related
FT         protein'
FT         146..1693
FT         mat_peptide
FT         /product='immunity related protein' FT CDS
FT         98..1696
FT         /product='immunity related protein' FT 3'UTR
FT         1697..1911.
FEATURES   Location/Qualifiers
           source
           1..1911
           /organism="Homo sapiens"
           /db_xref="taxon:9606"
BASE COUNT 490 a 541 c 525 g 355 t
ORIGIN

alignment_scores:
  Quality: 197.50      Length: 575
  Ratio: 0.906        Gaps: 20
  Percent Similarity: 36.000  Percent Identity: 18.957

alignment_block:
US-09-135-238b-2 x E15470 ..

Align seg 1/1 to: E15470 from: 1 to: 1911
a
3 ArgTrpLeuTrpProLeuTyPheLeuProValSerGlyAlaLeuArgI1 19
|||||
170 AGATGGCTGTGGGAG...GGCTCTCTCCCTCCAGGACCCATCTCCGGGC 216
|||||

19 eLeuProGluValLys..... 24
:::
217 CATGGGAACACTCAGGCCTTCTCGCCCTCTGCTGGGGGAGGAGAGCT 266
||:
25 .....ValGluGlyGlu 28
|||||
267 CCTTTGCAGCTCCAAATTCAATTGAAGGCTCAAGGCTGGTGTCAAGGGAG 316
|||||
29 LeuGlyGlySerValThrIleLysCysProLeuPro.....G1 41
|||||
317 CTGGAGGAGGTGTCACCATCGTGCATATATGCCCCCTCATCTGTCAA 366
|||||

41 uMethHisValArgIleTyLeuCysArgGluMetAlaGlySerGlyThrC 58
|||

```

```

: ||| ||| ||| ||| |||
367 CAGGCACACGAGGAAGTACTGGTGGCGTCTGGGGCCCCCAAGATGATCT 416
58 ySgLyThrValValSerThrThrAsnPheIleLysAlaGluTyLysGly 74
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
417 GCCAGACCATGTGTCCACCAACCAAGTATATCTACCATCGCTATCGTGAC 466
75 ArgValThrLeuLysGlnTyProArgLysAsnLeuPheLeuValGluVa 91
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
467 CGTGTGGCCCTCACAGACTTTCACAGAGAGAGGCTTGTGTGGTGAGGCT 516
91 lThrGlnLeuThrGluSerAspSerGlyValTyAlaCysGlyAlaGlyM 108
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
517 GTCCCAACTGTCCCGGATGACATCGGATGCTACCTCTCGCGCATTTGAA 566
108 etAsnThrAspArg..... 112
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
567 GTGAAAACAAACATGCTGTTCTTAAGCATGANTCTGACCATCTCTGCGAG 616
112 ..... 112
617 CCGGCCAGCACCTCCCCACAGCCACTCCAGCTGCTGGGAGCTCACCAT 666
113 .....GlyL 114
667 GAGATCCTATGGAACAGCGTCTCCAGTGGCCAAACAGATGGACCCCAAGAA 716
114 yThrGln..... 116
717 CCACCCAGACCTTAGACAGGGGACAGCATGGGACACAGTGTCTTCCACT 766
117 .....LysValThrLeuAsnValHisSerGluTyGluPro... 128
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
767 CCAGGAACCAAGCAAGACTACAGCTTCAGCTGAGGAAGACAGCAACCCAGG 816
129 .....SerTrpGluGlu..... 132
817 AGCAACCGCCAGCAGCTCCAGGGACAGCAGCTGGGCAGAGAGGTTCTG 866
133 .....GlnProMetProGluThrProLysTrpPheHisLeuPro 145
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
867 TCAAAGCAGCTGTCTCGGATTCCAGAGAGTCCA..... 898
146 TyrLeuPheGlnMetProAlaTyAlaSerSerLysPheValThrAr 162
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
899 .....COTTCAAAGACGAGAAGCATGTCCAA 924
162 gValThrThrProAlaGlnArgGlyLysValProProValHisHisSerS 179
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
925 TACAACAGAGAGGTGTGGAGGGC..... 949
179 erProThrThrGlnIleThrHisArgProArgValSerArg..... 192
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
950 ..ACCAGAAGCTCGGTGCACAAACAGGGCTAGAGCCAGCAAGGACAGGAGG 997
193 ...AlaSerSerValAlaGlyAspLysProArg..... 202
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
998 GAGATGACAACACTACCAAGGCTGATAGGCCAAGGAGGACATAGAGGGGT 1047
202 ..... 202
1048 CAGGATAGCTCTTGATGACGCCAAAAGGTCTTAGAACCATTTGGGCCAC 1097
203 .....ThrPheLeuProSerThrThr 209
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1098 CAGCTCTGGTCTCAGAAACTTTGGGCTGGGAAATCTCCCAACAAGCAAG 1147
210 AlaSerLysIleSerAlaLeuGluGlyLeuLeuLysProGlnThrProSe 226
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1148 CCAGTTTCTTAAGCAACAATCTCAGGGTTCCATTGGGAAACAACATCCAGC 1197
226 rTyAsnHisHisThrArg..... 232
: |||

```



```

1198 TGCAGCATGTGGACCTTGGAACTCCAGCTGCAGATGCTGGATCTTGG 1247
233 .....LeuHisArgGlnArgAlaLeuAspTyr 241
1248 GAACCTCCAGCTGCAGATGTGTGACAGCATGGAGCAGCATCTGGGAA 1297
242 GlySerGlnSerGly..... 246
1298 GGAAGCGCTGCAGGGACCTAGATGCTGCCACTGGACAGAGGTCCCCA 1347
246 ..... 246
1348 AGCAACACTGAGCCAGACCCCGCAGTAGAGACCTGGGGACCCCTGGCA 1397
247 .....ArgGluGlyGlnGlyPheHis 253
1398 AGGAGTCCTCGTGAAGCGTACTTTTCCAGAAGATGAAGACGACTCTCGG 1447
254 IleLeuIlePro.....ThrIleLeuGlyLeuPheLeuAlaLeuLe 268
1448 ACCTGCTCCTCTCTCTACCATGCTGGCCCTGTTTATGCTTATGGCTCT 1497
268 uGlyLeuValValLysArgAlaValGluArgGlyLeuAlaLeuSerArgA 285
1498 GGTTCATTATG.....CAAAGGAAGCTCTGGAGAAGGA 1529
285 rGlaArgArgLeuAlaValArgMetArgAlaLeuGluSerSerGln... 300
1530 GGACCTCTCAGAGCGCAAGAGGTCACTTAATTCAGATGACACATTTT 1579
301 .....ArgProArgGlySerProArgProArgSerGlnAsnIle 314
1580 CTGGAAGTGAACCCCAAGCAGACACAGCTGCCCCATGTGGAAAGAAAGAT 1629
314 eTyr.....SerAlaCysProArgArgAlaArgGlyAlaAspAlaAa 329
1630 GCTCCAGGATGACTCTCTTCT.....GCTGGGCCGACCC 1664
329 lyThrGlyGluAlaProValProGlyPro.....Gly 339
1665 TGACTGCCCCAGAGAAATCCAGGACCTGAGGACAGACAGATGAAC 1714
340 AlaProLeuPro.....ProAl 345
1715 GCTCAGTTACCTGGAGAGGACCAAGATCAAGGCTTCAGGACCCCA 1764
345 aProLeuGlnValSerGluSerPro 353
1765 GCCTCTTCCATCATCTCTCTCCA 1789

seq_name: gb_ro:RNIGR

seq_documentation_block: 3269 bp mRNA ROD 22-MAR-1995
LOCUS RNIGR Rat mRNA for polymeric immunoglobulin receptor.
DEFINITION X15741
ACCESSION X15741
VERSION X15741.1 GI:56464
KEYWORDS immunoglobulin receptor; polymeric immunoglobulin receptor; transmembrane protein.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 3269)
AUTHORS Banting,G.
TITLE Direct Submission
JOURNAL Submitted (05-JUL-1989) Banting G., EMBL, Meyerhofstr 1, 6900 Heidelberg
REFERENCE 2 (bases 1 to 3269)
AUTHORS Banting,G., Brake,B., Braghetta,P., Luzzio,J.P. and Stanley,K.K.
JOURNAL Unpublished
FEATURES
Location/Qualifiers
1..3269 source

```

```

/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/tissue_type="liver"
/clonelib="CDNA_TRL1"
/clonelib="GORF1,8,43,4339"
74..127
/product="rat polymeric immunoglobulin receptor signal peptide (AA -18 to -1)"
74..2383
/notes="rat polymeric immunoglobulin receptor (AA -18 to 751)"
/codon_start=1
/protein_id="CAA33758.1"
/db_xref="GI:56465"
/db_xref="SWISS-PROT:P15083"
/translation="MRLSLFALLVTVFSGVSTQSPFQPDVSSIEGNSVSTCYYPDTSVNRHTRKYWCROGANGCATLISNGYLSKEYSGRASLINFENSTFVNIHLTQEDTGYKGLGTNRGLPFDVLEVSQVPEFNDTHVTKDTRGTITIECRKEGNAH SKSLACKRGEACEVVIDSTEYVDPYKDRALLFMKGTSDRDIYVNIHLIPSDAGLY VCQAGESPADKNADLOVLEPELLELYKDLSSVTECDLGRVANDAKYLCKRKE TCDVILNLGKRDPAPFGRILLIPRODNGRFSVLITLKRDEAGHYCQGAHSSGLPQE GWPQAWQLFVNEESTIPNSRSVVKGVGTGSVAIVCPYNKRESSLKYWCHEADENG RCPVLVQALVOEGYEGRLADPQGSAYTVILNLTQDSFGYVCLTDGDSKWT TIELOVAEATKKPDLVTPQNAIVIGETFTISCHPCFKYQKWKWNSDGHIL PSHDEGARSSVCDQSQSVIATNPVKKEDEGWYKGVKGYGETTATYVAVEE RTGSGPHINTDANARAKDAPPEEAMESSVREDENKNDLPRLFADERIONAGDQAO ENRASNAGSAGSGSSKVLFTLVPLGLVAVGAVAVVAVRVHRKKNVDRMSISSY RTDISMGDFNSRDLGNGNMGATPQETVLEKDELETTTCTTEPEESKKARRS KERADMAYSANFLFQSTTAQVHDGPQA"
128..2380
mat_peptide /product="rat polymeric immunoglobulin receptor mature peptide (AA 1 to 751)"
BASE COUNT 890 a 831 c 850 g 698 t
ORIGIN

alignment_scores:
Quality: 185.50 Length: 164
Ratio: 1.855 Gaps: 6
Percent Similarity: 60.976 Percent Identity: 30.488

alignment_block:
US-09-135-238b-2 x RNIGR ..
Align seg 1/1 to: RNIGR from: 1 to: 3269

30 GlyGlySerValThrIleLysCysProLeuProGlu.....Me 42
|||||:|||||:|||||:|||||:
170 GGTAACCTCGGTCTCCATCAGCTGCTACTACCAGACACCTCTGTCAACCG 219

42 thisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 59
|||||:|||||:|||||:|||||:
220 GCACACCCGGAAATATCTGTCGCCACAGGAGCC...AACGGCTACTGCG 266

59 lyThrValValSerThrAsnPhelIleLysAlaGluTyrLysGlyArg 75
:::|||||:|||||:|||||:|||||:
267 CAACCTCATCTCTTCAATGGCTACCTCTCGAAGGAGTATTCAGGCAGA 316

76 ValThrLeuLysGlnTyrProArgLysAsnPhelLeuValGluValTh 92
:::|||||:|||||:|||||:|||||:
317 GCACCTCATCACTCACTCCAGAGATGACATTTGTGATTACATGCG 366

92 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMetA 109
:::|||||:|||||:|||||:|||||:
367 ACATCTCACCCAGGAGGACACTGGGAGCTACAAAGTGTGCTGGGT...A 413

109 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 125
:::|||||:|||||:|||||:|||||:
414 CCACTAACCCAGGCGCTGTTTTTCGATGTCAGCCTGGAGGTC..... 454

126 TyrGluProSerTrpGluGlnProMetProGluThrProLysTrpPh 142
:::|||||:|||||:|||||:
455 .....AGCCAGGTCTCTGAGTTCCTCCAAATGACAC 483

```

142 ehIsLeuProTyLeuPheGlnMetProAlaTyAlaSerSerSerLysp 159  
|||||  
484 CCATGTC.....TACACAAAGG 500  
  
159 heValThrArgValThrProAlaGlnArgGlyLysValProProVal 175  
::: |||:||||| || |||  
501 ACATAGGCAGAACTGTGACATCGAATGCCGTTTCAAGAGGGGAATGCT 550  
  
176 HisHisSer.SerProThrThrGlnIleThrHisArgPro 188  
||| ::::|::| ::::|::|  
551 CATACGAGAATCCCTGTGTGAAGAGAGAGAGAGGCCT 590

seq\_name: gb\_om:AB032195

seq\_documentation\_block: 2461 bp mRNA MAM 03-FEB-2000  
LOCUS AB032195  
DEFINITION Sus scrofa mRNA for poly-Ig receptor, complete cds.  
ACCESSION AB032195  
VERSION AB032195.2 GI:6863079  
KEYWORDS poly-Ig receptor.  
SOURCE Sus scrofa female cDNA to mRNA, clone:SC1.  
ORGANISM Sus scrofa  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
Sone.T. and Kumura.H.  
TITLE Porcine mammary gland cDNA clone, similar to poly-Ig receptor  
JOURNAL Published Only in DataBase (1999) In press  
REFERENCE 2 (bases 1 to 2461)  
AUTHORS Sone.T. and Kumura.H.  
TITLE Direct Submission  
JOURNAL Submitted (07-SEP-1999) to the DBJ/EMBL/GenBank databases. Teruo  
Sone, Hokkaido University, Laboratory of Applied Microbiology,  
Faculty of Agriculture, Kita-9, Nishi-9, Kita-ku, Sapporo, Hokkaido  
060-8589, Japan (E-mail:sonet@chem.agr.hokudai.ac.jp,  
Tel:81-11-706-2502, Fax:81-11-706-4961)  
COMMENT On Feb 3, 2000 this sequence version replaced gi:5881588.  
Sequence updated (01-Feb-2000).

FEATURES  
Location/Qualifiers  
1..2461  
/organism="Sus scrofa"  
/db\_xref="taxon:9823"  
/clone="SC1"  
/sex="female"  
58..111  
/product="putative"  
58..2334  
/codon\_start=1  
/product="poly-Ig receptor"  
/protein\_id="BAA84283.2"  
/db\_xref="GI:6863080"  
/translation="MTRFFYACLLAIFPVVSMKSPITFGPDVSVSGSVSRICYPA  
TSVNRHSRYWCIRIGAGKRCCTLLISSEGYISKDYGRANLTNFPENGTFVMDIHLTR  
GDSGLYKCLGILSSRGLSDVLEVSQPGQIGDVHYTADLGSTVTINCPKPSNAQ  
KPSVYKGLQIRVLVDSNGYLNNTNRAHLSIQGTNOLVFSVINRIOLRKGDIY  
ICQAGDESSADQLVLPPELYIGDLRGSVTFDCALGOEMANVAKFLCOLKNGKTCN  
VVTNTLGAQVPEGRILLTPKENSFSVHTLGRKEDAGHYLCGHPDGFKEGWPV  
QANQLINETHMIPRPSVVRGVVGSVATCIPNPKETNSLKWREENENGRCPQ  
LVSSGLVNEQYGRALYEEPEGNGFTTILNQLTNRDAGFYWCLTNEDSRWSTVEL  
KIVEPEPNLKPENYAWGETLKLSCFPCKFYQYKWKWSNTGCRALPSQDEGQ  
SQAVNCDDKKSQIISLNLPVRKEDEGWKCVKGLHYGETGAVTVVAVQKAKGSGD  
ARUANAPAPAEADIEPRARETENEVLDPSLIAEDRAVKDGEPSAGSGVPAAPGSS  
VGQGGSKAVSTLPLALVMLVGAVTIGVLRARHKVNDRLSISRTDLSMSDFES  
SRDFGHDNAGATLDQETLSLGGKDRFTTTTENTMEDREKRAKRSKEEADAKATTFI  
LQANSIIIAQTONGPREA"  
BASE COUNT 608 a 699 c 708 g 446 t  
ORIGIN

alignment\_scores:  
Quality: 184.00 Length: 120  
Ratio: 2.217 Gaps: 4

Percent Similarity: 69.167 Percent Identity: 34.167  
alignment\_block:  
US-09-135-238B-2 x AB032195 ..  
  
Align seg 1/1 to: AB032195 from: 1 to: 2461  
  
17 LeuArgIleLeuProGluValLysValGluGlyGluLeu..... 29  
||| |||:||||| |||:|||||:|::|  
85 CTGGCTATCTTCCCACTGGTCTCCATGAAGAGTCCCATATTCGGTCCCA 134  
|||:|||||:|::|  
30 .....GlyGlySerValThrIleLysCysProLeuP 40  
|||:|||||:|::|  
135 GGATGTGAGCAGCGTGAAGGAGCTCGGTGTCCATCAGATGCTACTACC 184  
  
40 ro.....GluMetHisValArgIleTyrLeuCysArgGluMet 52  
|| ::::|::| ||| ||| ||| |||  
185 GACCCACCTCCGTCAACCGGCATCTCGGAAGTACTGGTCCCA...ATA 231  
:  
53 AlaGlySerGlyThrCysGlyThrValValSerThrThrAsnPheIleLe 69  
:|||||:|::| ||| |||:|||||:|::|  
232 GGAGCCAAAGGCGCGTCACACCCCTCATCTCTCGGAGGCTACATCTC 281  
:  
69 sAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProArgLysAsnL 86  
: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
282 CAAGGACTACAAGGCGAGAGCCCAACCTCACCACACTCCACAGAACGGCA 331  
  
86 euPheLeuValGluValThrGlnLeuThrGluSerAspSerGlyValTyr 102  
|||:|||||:|::| ||| |||:|||||:|::| ||||| ||||| |||||  
332 CCTTCGTGATGGACATTGGCCACCTGACCGCGGTGACTCTGGGCTCTAC 381  
  
103 AlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLysValTh 119  
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
382 AAGTGGGCTGGGCATTAGCAGC...CGAGGCCTCTCTTTGATGTGAG 428  
  
119 rLeuAsnVal 122  
:|||||:|::|  
429 CCTGGAGGTC 438

seq\_name: gb\_om:BOVPIGR

seq\_documentation\_block:  
LOCUS BOVPIGR 3527 bp mRNA MAM 30-JAN-1995  
DEFINITION Bos taurus polymeric immunoglobulin receptor mRNA, complete cds.  
ACCESSION L04797  
VERSION L04797.1 GI:388279  
KEYWORDS immunoglobulin receptor; polymeric immunoglobulin receptor.  
SOURCE Bos taurus (tissue library: lambda gt10) small intestine cDNA to mRNA.

ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovinae; Bos.  
REFERENCE 1 (bases 1 to 3527)  
AUTHORS Kulseth,M.A., Krajci,P., Myklebost,O. and Rognes,S.  
TITLE Cloning and characterization of two forms of bovine polymeric  
immunoglobulin receptor cDNA  
JOURNAL DNA Cell Biol. 14 (3), 251-256 (1995)  
MEDLINE 95186063  
FEATURES  
Location/Qualifiers  
1..3527  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/tissue\_type="small intestine"  
/tissue\_lib="lambda gt10"  
73..2346  
/standard\_name="pIgR"  
/codon\_start=1  
/product="polymeric immunoglobulin receptor"  
/protein\_id="AAC41620.1"  
/db\_xref="GI:388280"  
/translation="MSRLFLACLIAIFPVVSMKSPITFGPEVTSVEGRSVSIKCYVPP  
TSVNRHTRKYWCROGAQRCTTLLISSGYVDDYVGRANLTNFPESGTFVVDISHLTH



LOCUS S43435S03 345 bp DNA PRI 08-MAY-1993  
DEFINITION transmembrane secretory component [human, leukocytes, Genomic, 345  
nt, segment 3 of 11].  
ACCESSION S43441  
VERSION S43441.1 GI:255089  
KEYWORDS 3 of 11  
SOURCE human leukocytes.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;  
1 (bases 1 to 345) Eutheria; Primates; Catarrhini; Homidae; Homo.  
AUTHORS Krajci,P., Kvale,D., Tasken,K. and Brandtzaeg,P.  
TITLE Molecular cloning and exon-intron mapping of the gene encoding  
human transmembrane secretory component (the poly-Ig receptor)  
JOURNAL Eur. J. Immunol. 22 (9), 2309-2315 (1992)  
MEDLINE 92387236  
REMARK GenBank staff at the National Library of Medicine created this  
entry [NCBI gibbsq 113232] from the original journal article.  
This sequence comes from Fig. 2.

FEATURES  
source Location/Qualifiers  
1..345  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 78 a 102 c 97 g 68 t  
ORIGIN

alignment\_scores:  
Quality: 182.00 Length: 115  
Ratio: 2.247 Gaps: 5  
Percent Similarity: 70.435 Percent Identity: 37.391

alignment\_block:  
US-09-135-238B-2 x S43435S03 ..  
Align seg 1/1 to: S43435S03 from: 1 to: 345

12 ProValSerGlyAlaLeuArgIleLeuProGluValLysValGluGly 28  
||||| ||| ||||| :|||:|||||  
18 CCATATTGGT.....CCCGAG...GAGGTGAATAGTGT 49  
28 uLeuGlyGlySerValThrIleLysCysProLeuPro.....G 41  
|||||:|||||: ||| |||||: ||| :|||  
50 GGAAGGTAACCTCAGTGTCCATCAGCTGCTACTACCCACCCCTCTGTCA 99  
41 LuMetHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThr 57  
:: |||:|||| ||| |||||: ||| :|||  
100 ACCGGCACACCCGGAAGTACTGGTCCGCGGAGGAGCTAGAGGTGGC... 146

58 CysGlyThrValValSerThrThrAsnPheIleLysAlaGluTyrLysG1 74  
||| |||:|||||: ||| :|||:|||||: ||| |||  
147 TGCATAACCTCATCTCTCGGAGGGCTACGTCCTCCAGCAATATGCAGG 196

74 YArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluV 91  
|||||:|||||: |||:|||||: |||:|||||: |||:|||||: |||  
197 CAGGGCTAACCTCACCAACTTCCCGAGAACGGCACATTTCGTGGTGAACA 246

91 alThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGly 107  
::|||:|||||: ||| |||||: ||| ||||| |||  
247 TTGCCAGCTGAGCCAGGATGACTCCGGCGGCTACAAGTGTGGCTGGGC 296

108 MetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnVal 122  
::|||:|||||: |||: |||: |||: |||: |||: |||: |||  
297 ATCATATAGC...CGAGGCTGTCCTTTGATGTCAGCCTGGAGGTC 338

seq\_name: gb\_pat:A52094

seq\_documentation\_block:  
LOCUS A52094 1936 bp DNA PAT 11-MAR-1997  
DEFINITION Sequence 4 from Patent WO9618734.  
ACCESSION A52094  
VERSION A52094.1 GI:2304706

KEYWORDS .  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 1936)  
AUTHORS Van,O.J., Asselbergs,F.A., Leist,C. and Gruetter,M.  
TITLE PRODUCTION OF RECOMBINANT SECRETORY COMPONENT  
JOURNAL Patent: WO 9618734-A 4 20-JUN-1996;  
CIBA GEIGY AG (CH)  
COMMENT Other publication AU 4302796 960703.  
FEATURES Location/Qualifiers  
1..1936  
/organism="unidentified"  
/db\_xref="taxon:32644"  
16..1920  
mat\_peptide /product="TRANSCRIPT FROM SC DNA FRAGMENT 2"  
BASE COUNT 451 a 521 c 583 g 381 t  
ORIGIN

alignment\_scores:  
Quality: 182.00 Length: 115  
Ratio: 2.247 Gaps: 5  
Percent Similarity: 70.435 Percent Identity: 37.391

alignment\_block:  
US-09-135-238B-2 x A52094 ..  
Align seg 1/1 to: A52094 from: 1 to: 1936

12 ProValSerGlyAlaLeuArgIleLeuProGluValLysValGluGly 28  
||||| ||| ||||| :|||:|||||  
76 CCATATTGGT.....CCCGAG...GAGGTGAATAGTGT 107

28 uLeuGlyGlySerValThrIleLysCysProLeuPro.....G 41  
|||||:|||||: ||| |||||: ||| :|||  
108 GGAAGGTAACCTCAGTGTCCATCAGCTGCTACTACCCACCCCTCTGTCA 157

41 LuMetHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThr 57  
:: |||:||||| ||| |||||: ||| :|||  
158 ACCGGCACACCCGGAAGTACTGGTCCGCGGAGGAGCTAGAGGTGGC... 204

58 CysGlyThrValValSerThrThrAsnPheIleLysAlaGluTyrLysG1 74  
||| |||:|||||: ||| :|||:|||||: ||| |||  
205 TGCATAACCTCATCTCTCGGAGGGCTACGTCCTCCAGCAATATGCAGG 254

74 YArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluV 91  
|||||:|||||: |||:|||||: |||:|||||: |||:|||||: |||  
255 CAGGGCTAACCTCACCAACTTCCCGAGAACGGCACATTTCGTGGTGAACA 304

91 alThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGly 107  
::|||:|||||: ||| |||||: ||| ||||| |||  
305 TTGCCAGCTGAGCCAGGATGACTCCGGCGGCTACAAGTGTGGCCTGGGC 354

108 MetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnVal 122  
::|||:|||||: |||: |||: |||: |||: |||: |||: |||  
355 ATCAATAGC...CGAGGCTGTCCTTTGATGTCAGCCTGGAGGTC 396

---

OM of: US-09-135-238B-2 to: Issued Patents\_NA.\* out\_format : pfs  
 Date: Sep 12, 2000 5:38 AM  
 About: Results were produced by the GenCore software, version 4.5,  
 Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters: -MODEL-frame-p2n.model -DBV-x1p  
 -Q/cgn2\_1/USPTO.spool/US0915238/runat\_24082000\_091114\_16791/app\_query.fasta\_1.1282  
 -DB-Issued Patents\_NA -QFMT-fastap -SUFFIX=p2n.rni  
 -GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000  
 -LOOPEXT=0.000 -OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000  
 -XGAPEXT=0.500 -DLOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000  
 -MATRIX=bloms62 -TRANS=human40.cdi -LIST=45 -DALIGN=200  
 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
 -OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=100000  
 -USER=US0915238\_CG1\_1\_68 -NCPU=6 -ICPU=3 -LONGLOG -NO\_XLXPY  
 -WAIT -THREADS=1

Search information block:  
 Query: US-09-135-238B-2  
 Query length: 390  
 Database: Issued Patents\_NA.\*  
 Database sequences: 243080  
 Database length: 68777915  
 Search time (sec): 94.120000

score_list:	Seq	Strd	Orig	2Score	EScore	Len	Document	1
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-434-000A-7	!	188.50	262.73	3.8e-07	3095	!		
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-434-000A-9	!	185.50	243.58	4.5e-06	3269	!		
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-434-000A-3	!	182.00	239.68	7.4e-06	2919	!		
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-434-000A-5	!	182.00	237.20	1.0e-05	3630	!		
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-642-406A-21	!	166.50	221.76	7.3e-05	1875	!		
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-642-406A-20	!	166.50	214.59	0.0002	3517	!		
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-434-000A-1	!	166.50	214.59	0.0002	3517	!		
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-935-450-10	!	142.00	193.65	0.0027	914	!		
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-824-405-5	!	136.50	178.54	0.0187	1683	!		
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-444-644-27	!	135.00	155.12	0.3776	10785	!		
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-444-644-32	!	131.00	147.85	0.9588	12127	!		
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-378-939-9	!	130.00	169.36	0.0608	1617	!		
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-444-644-18	!	129.50	146.21	1.18	11528	!		
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-157-101A-6	!	125.50	162.98	0.1377	1576	!		
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-713-118-1	!	125.50	145.55	1.29	7266	!		
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-286-740-3	!	124.50	145.24	1.34	6557	!		
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-286-740-3	!	124.50	145.24	1.34	6557	!		
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-576-626A-2	!	123.00	140.67	2.41	8051	!		
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-149-099C-3	!	123.00	139.14	2.93	9209	!		
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-478-967A-3	!	123.00	139.14	2.93	9209	!		
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-819-866-2	!	123.00	130.88	8.45	18986	!		
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-09-023-715-2	!	123.00	130.88	8.45	18986	!		
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-480-036-1	!	122.50	154.29	0.4200	2287	!		
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-461-968A-1	!	122.50	154.29	0.4200	2287	!		
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-462-571-1	!	122.50	154.29	0.4200	2287	!		
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-462-571-1	!	122.50	154.29	0.4200	2287	!		
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-916-917-1	!	122.50	146.09	1.20	4692	!		
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-973-631-1	!	122.50	146.09	1.20	4692	!		
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-973-631-1	!	122.50	146.09	1.20	4692	!		
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-973-631-1	!	122.50	146.09	1.20	4692	!		
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-672-211-1	!	122.50	146.09	1.20	4692	!		
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-09-225-170-1	!	122.50	146.09	1.20	4692	!		
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-455-543A-7	!	121.50	139.47	2.81	7362	!		
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-193-078B-7	!	121.50	139.47	2.81	7362	!		
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-223-305C-7	!	121.50	139.47	2.81	7362	!		
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-149-097D-7	!	121.50	139.47	2.81	7362	!		
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-310-370-1	!	121.00	159.52	0.2147	1190	!		
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-955-937A-1	!	119.50	149.56	0.7705	2345	!		
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-444-644-41	!	119.50	132.09	7.24	10844	!		

/cgn2\_6/ptodata/1/ina/5B\_COMB.seq:US-08-764-233A-1 - 118.50 113.31 80.51 49377  
 /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:US-08-455-543A-8 + 118.00 134.57 5.26 7175  
 /cgn2\_6/ptodata/1/ina/5C\_COMB.seq:US-08-193-078B-8 + 118.00 134.57 5.26 7175  
 /cgn2\_6/ptodata/1/ina/5C\_COMB.seq:US-08-223-305C-8 + 118.00 134.57 5.26 7175  
 /cgn2\_6/ptodata/1/ina/5C\_COMB.seq:US-08-149-097D-8 + 118.00 134.57 5.26 7175

seq\_name: /cgn2\_6/ptodata/1/ina/6\_COMB.seq:US-08-434-000A-7

seq\_documentation\_block:  
 ; Sequence 7, Application US/08434000A  
 ; Patent No. 6046037  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ANDREW C. HIATT, JULIAN  
 ; APPLICANT: K.C. MA, THOMAS LEHNER  
 ; TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION  
 ; NUMBER OF SEQUENCES: 19  
 ; TITLE OF INVENTION: PROTEINS IN PLANTS AND THEIR USES  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: LYON & LYON  
 ; STREET: 633 West Fifth Street  
 ; STREET: Suite 4700  
 ; CITY: Los Angeles  
 ; STATE: California  
 ; COUNTRY: U.S.A.  
 ; ZIP: 90071  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
 ; MEDIUM TYPE: storage  
 ; OPERATING SYSTEM: IBM P.C. DOS 5.0  
 ; SOFTWARE: Word Perfect 5.1  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/434,000A  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; PRIOR APPLICATION DATA: including application  
 ; PRIOR APPLICATION DATA: described below: 1  
 ; APPLICATION NUMBER: 08/367,395  
 ; FILING DATE: 12/30/94  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Guise, Jeffrey W.  
 ; REGISTRATION NUMBER: 34,613  
 ; REFERENCE/DOCKET NUMBER: 212/127  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (619) 552-8400  
 ; TELEFAX: (619) 552-0159  
 ; TELEX: 67-3510  
 ; INFORMATION FOR SEQ ID NO: 7:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3095 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; FEATURE: DESCRIPTION: Mouse Polyimmunoglobulin Receptor  
 ; NAME/KEY: Coding Sequence  
 ; LOCATION: 85....2400  
 ; US-08-434-000A-7

alignment\_scores:  
 Quality: 198.00 Length: 168  
 Ratio: 1.980 Gaps: 5  
 Percent Similarity: 59.524 Percent Identity: 31.548

alignment\_block:  
 US-09-135-238B-2 x US-08-434-000A-7 ..

Align seg 1/1 to: US-08-434-000A-7 from: 1 to: 3095

10 PheLeuProValSerGlyAlaLeuArgIleuProGluValIysValIgl 26

```

121 TTTTCAGGGGCTCCACAAAGCCCATATTTGGTCCCGAGGAGTGAG 170
122 TTTTCAGGGGCTCCACAAAGCCCATATTTGGTCCCGAGGAGTGAG 170
26 uGlyGluLeuGlySerValThrIleLysCysProLeuProGlu... 41
27 TTTTCAGGGGCTCCACAAAGCCCATATTTGGTCCCGAGGAGTGAG 170
171 TAGTATAGAGGAGCTCTGTTTCCATCAGCTGTACTACCCAGACACT 220
172 TTTTCAGGGGCTCCACAAAGCCCATATTTGGTCCCGAGGAGTGAG 170
42 .....MetHisValArgIleTyrLeuCysArgGluMetAlaGlySer 55
43 TTTTCAGGGGCTCCACAAAGCCCATATTTGGTCCCGAGGAGTGAG 170
221 CTGTCAACCGGCACACCCGAAATACATGTCGCGACAGAGGCC...AGC 267
222 TTTTCAGGGGCTCCACAAAGCCCATATTTGGTCCCGAGGAGTGAG 170
56 GlyThrCysGlyThrValValSerThrThrAsnPhelLysAlaGly 72
57 TTTTCAGGGGCTCCACAAAGCCCATATTTGGTCCCGAGGAGTGAG 170
268 GCATGTCGCAACAGCTCATCTCTCAATGGCTACTCTCCAAAGAGTA 317
269 TTTTCAGGGGCTCCACAAAGCCCATATTTGGTCCCGAGGAGTGAG 170
72 rlyGlyArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeu 89
73 TTTTCAGGGGCTCCACAAAGCCCATATTTGGTCCCGAGGAGTGAG 170
318 TTTTCAGGGGCTCCACAAAGCCCATATTTGGTCCCGAGGAGTGAG 170
89 aGluValThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGly 105
90 TTTTCAGGGGCTCCACAAAGCCCATATTTGGTCCCGAGGAGTGAG 170
368 TTAACATTGAGCAGCTACCCAGGAGCAGACTGGGAGCTACAAGTGTG 417
369 TTTTCAGGGGCTCCACAAAGCCCATATTTGGTCCCGAGGAGTGAG 170
106 AlaGlyMetAsnThrAspArgGlyThrGlnLysValThrLeuAsnVa 122
107 TTTTCAGGGGCTCCACAAAGCCCATATTTGGTCCCGAGGAGTGAG 170
418 CTGGGT...ACCAGTAACGAGGCTCTCTTCGATGTCAGCTGGAGT 464
419 TTTTCAGGGGCTCCACAAAGCCCATATTTGGTCCCGAGGAGTGAG 170
122 lHisSerGluTyrGluProSerTyrGluGlnProMetProGluThr 139
123 TTTTCAGGGGCTCCACAAAGCCCATATTTGGTCCCGAGGAGTGAG 170
465 C.....AGCCAGTCTCTGAGTTCG 484
466 TTTTCAGGGGCTCCACAAAGCCCATATTTGGTCCCGAGGAGTGAG 170
139 rlyTyrTrpPheHisLeuProTyrLeuPheGlnMetProAlaTyrAla 155
140 TTTTCAGGGGCTCCACAAAGCCCATATTTGGTCCCGAGGAGTGAG 170
485 CGAGTGACACCCAGCTC.....TACACAAAG 510
486 TTTTCAGGGGCTCCACAAAGCCCATATTTGGTCCCGAGGAGTGAG 170
156 SerSerLysPheValThrArgValThrProAlaGlnArgGlyLysVa 172
157 TTTTCAGGGGCTCCACAAAGCCCATATTTGGTCCCGAGGAGTGAG 170
511 GACATAGCAGAAATGACCATTTGATGCTTTCAAAGGAGGAGTGT 560
512 TTTTCAGGGGCTCCACAAAGCCCATATTTGGTCCCGAGGAGTGAG 170
172 lPro 173
173 TTTTCAGGGGCTCCACAAAGCCCATATTTGGTCCCGAGGAGTGAG 170
561 TCCT 564
562 TTTTCAGGGGCTCCACAAAGCCCATATTTGGTCCCGAGGAGTGAG 170

```

```

seq_name: /cgn2_6/ptodata/1/ina/6_5_comb.seq:us-08-434-000A-9

```

```

seq documentation block:
Sequence 9, Application US/08434000A
Patent No. 6046037
GENERAL INFORMATION:
APPLICANT: ANDREW C. HIATT, JULIAN
TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
TITLE OF INVENTION: PROTEINS IN PLANTS AND THEIR USES
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/434,000A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

```

```

PRIOR APPLICATION DATA: including application
PRIOR APPLICATION DATA: described below:
APPLICATION NUMBER: 08/367,395
FILING DATE: 12/30/94
ATTORNEY/AGENT INFORMATION:
NAME: Guise, Jeffrey W.
REGISTRATION NUMBER: 34,613
REFERENCE/DOCKET NUMBER: 212/127
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 552-8400
TELEFAX: (619) 552-0159
TELEX: 67-3510
TELEX: SEQUENCE LISTING
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 3269 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE: DESCRIPTION: Rat Polyimmunoglobulin Receptor
NAME/KEY: Coding Sequence
LOCATION: 74....2383
US-08-434-000A-9

alignment_scores:
Quality: 185.50 Length: 164
Ratio: 1.855 Gaps: 6
Percent Similarity: 60.976 Percent Identity: 30.488

alignment_block:
US-09-135-238B-2 x US-08-434-000A-9 ..
Align seq 1/1 to: US-08-434-000A-9 from: 1 to: 3269
30 GlyGlySerValThrIleLysCysProLeuProGlu.....Me 42
170 GGTAACTCGTCTCCATCAGCTGCTACTACCCAGACACCTCTGTCAACGG 219
42 tHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 59
220 GCACACCCGGAATCTGCTGCCGACAGAGGCC...AACGGCTACTGGG 266
59 lYThrValValSerThrThrAsnPhelLysAlaGluTyrLysGlyArg 75
267 CAACCTCTCTCTCAATGCTACTCTCCGAGGAGTATCAGCAGA 316
76 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluVal 92
317 GCACGCTCATCACTACCTCCAGAGAAATAGCATTGTGATTAACATGCG 366
92 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 109
367 ACATCTCACCCAGGAGGACACTGGGAGCTACAAAGTGTGGTGGGT...A 413
109 snThrAspArgGlyThrGlnLysValThrLeuAsnValHisSerGlu 125
414 CCACCTAACCGAGGCGCTGTTTTCGATGTCAGCTGGAGGTC..... 454
126 TyrGluProSerIrrpGluGluInProMetProGluThrProLysTrp 142
455 .....AGCCAGTCTCTGAGTTCCTCCAAATGACAC 483
142 eHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSerSerSerLysP 159
484 CCAATGTC.....TACACAAAGG 500
159 heValThrArgValThrThrProAlaGlnArgGlyLysValProVal 175
501 ACATAGGAGAACTGTGACCATTCGAATGCCGTTTCAAAGAGGAGGAGTGT 550
176 HisHisSer.SerProThrThrGlnIleThrHisArgPro 188

```





```

; TOPOLOGY: DESCRIPTION: Bovine Polymunoglobulin Receptor
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 152.....2425
; US-08-434-000A-5

alignment_scores:
  Quality: 182.00      Length: 123
  Ratio: 2.193        Gaps: 4
  Percent Similarity: 57.480      Percent Identity: 34.146

alignment_block:
  US-09-135-238B-2 x US-08-434-000A-5 ..
  Align seg 1/1 to: US-08-434-000A-5 from: 1 to: 3630

17 LeuArgileLeuProGluValIysValIgluGlyGluLeu..... 29
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
179 CTGGCCATCTTCCAGTGTCTCCATGAAGAGTCCCATCTTCGGTCCCGA 228
30 .....GlyGlySerValThrIleLysCysProLeuP 40
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
229 GGAGGTGAGCAGCGTGGAGCGCGCTCAGTGTCCATCAGTGTACTACC 278
40 ro.....GluMetHisValArgIleTyrLeuCysArgGluMet 52
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
279 CGCCACCTCCGTCACCGCGCACGCGCAAGTACTGTGCGCGCAGGGA 328
53 AlaGlySerGlyThrCysGlyThrValSerThrThrAsnPheIlely 69
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
329 GCC...CAGGCGCGCTGCACGACCTCATCTCTCGAGGCGTACGTCTC 375
69 sAlaGlyTyrIysGlyArgValThrLeuLysGlnTyrProArgIysAsnL 86
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
376 CGAGGACTACTGGCGCAGAGCAACCTCACCACCTCCCGAGAGCGGCA 425
86 eupheLeuValGluValThrGlnLeuThrGluSerAspSerGlyValTyr 102
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
426 CGTTGTGTGGATCAGCAGCATCTCACCATAAGACTCAGGCGCTAC 475
103 AlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLysValth 119
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
476 AGTGTGCGCTGGCATTTAGCAGC...CGTGGCTTAATCTCGATGTGAG 522
119 rIeuAsnValHisSerGlu 125
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
523 CTGGAGGTACGCCAAGAT 541

seq_name: /cgn2_5/ptodata/1/ina/5D_COMB.seq:US-08-642-406A-21
seq_documentation_block:
; Sequence 21, Application US/08642406A
; Patent No. 5959177
; GENERAL INFORMATION:
; APPLICANT: Heintz, Mich B.
; APPLICANT: Hiatt, Andrew C.
; APPLICANT: Ma, Julian K.C.
; TITLE OF INVENTION: TRANSGENIC PLANTS EXPRESSING ASSEMBLED
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THE SCRIPPS RESEARCH INSTITUTE
; STREET: 10666 No. 5959177th Torrey Pines Road, TPC-8
; CITY: La Jolla
; STATE: California
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642,406A
; FILING DATE: 03-MAY-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/591,823
; FILING DATE: 02-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/427,765
; FILING DATE: 27-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Logan, April C.
; REGISTRATION NUMBER: 33,950
; REFERENCE/DOCKET NUMBER: 184.2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 554-2837
; TELEFAX: (619) 554-6312
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1875 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1875
; US-08-642-406A-21

alignment_scores:
  Quality: 166.50      Length: 103
  Ratio: 2.602        Gaps: 4
  Percent Similarity: 62.136      Percent Identity: 42.718

alignment_block:
  US-09-135-238B-2 x US-08-642-406A-21 ..
  Align seg 1/1 to: US-08-642-406A-21 from: 1 to: 1875

30 GlyCysSerValThrIleLysCysProLeuPro.....GluMe 42
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
115 GGGGACTCGTGTCTCCATCAGTGTCTACTACCAACCACTCCGTCACCCG 164
42 tHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 59
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
165 GCACAGCCGGAAGTCTGTGTGCGGGAA...GAGGAGAGCGCGCTGCG 211
59 lyThrValSerThrThrAsnPheIleLysAlaGlyTyrLysGlyArg 75
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
212 TGAGCGTTGCTCG...ACGGCTACAGTCCCGCAATCTCCGCGAGA 258
76 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValth 92
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
259 GGCAGCTCAGCGACTTCCCTGATAGGGGAGTGTGGTGGTGTGTA 308
92 rGlnLeuThrGluSerSerGlyValTyrAlaCysGlyAlaGlyMetA 109
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
309 CCACACACCCAGACGACTCAGGAGGAGTCAAGTGTGCGTGGGAGTCA 358
109 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 125
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
359 AC...GGCGGTGGCTGGCTCGGTGTCACAGTGTGTCAGCGCAGAG 405
126 TyrGluPro 128
||| |||||
406 CCAGAGCGCT 414

seq_name: /cgn2_5/ptodata/1/ina/5D_COMB.seq:US-08-642-406A-20
seq_documentation_block:

```

```

; Sequence 20, Application US/08642406A
; Patent No. 5959177
; GENERAL INFORMATION:
; APPLICANT: Hein, Mich B.
; APPLICANT: Hiatt, Andrew C.
; APPLICANT: Ma, Julian K.C.
; TITLE OF INVENTION: TRANSGENIC PLANTS EXPRESSING ASSEMBLED
; SECRETORY ANTIBODIES
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THE SCRIPPS RESEARCH INSTITUTE
; STREET: 10666 No. 595917th Torrey Pines Road, TPC-8
; CITY: La Jolla
; STATE: California
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642,406A
; FILING DATE: 03-MAY-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/591,823
; FILING DATE: 02-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/427,765
; FILING DATE: 27-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Logan, April C.
; REGISTRATION NUMBER: 33,950
; REFERENCE/DOCKET NUMBER: 184.2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 554-2937
; TELEFAX: (619) 554-6312
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3517 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-642-406A-20

alignment_scores:
  Quality: 166.50      Length: 103
  Ratio: 2.602        Gaps: 4
  Percent Similarity: 62.136      Percent Identity: 42.718

alignment_block:
  US-09-135-238B-2 x US-08-642-406A-20  ..

Align seg 1/1 to: US-08-642-406A-20 from: 1 to: 3517

30 GlyGlySerValThrIleLysCysProLeuPro.....Glume 42
   ||| |||||:||||| ||| |||
238 GCGACGCTGGTGCATCACATGCTACTACCCACACCTCGTCACCG 287
42 tHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 59
   ||| ||| :||| ||||| |||
288 CCACAGCGCGAGTTCGTGGTGGCGGAA...GAGGAGAGCGCGCGTCG 334
59 lYThrValValSerThrThrAsnPhelIleLysAlaGluTyrLysGlyArg 75
   |||:||||| ||| |||||: ||| |||||: |||||
335 TCAGCGCTGGCTCG...ACCGGTACACGTCGCCAGGAATCTCCGGGAGA 381
76 ValThrLeuLysGlnTyrProArgLysAsnPhelLeuValGluValThr 92

```

```

||| :||||| |||: |||: |||: |||
382 GCGAAGCTCACCAGTCCCTGATAAAGGGAGTTGTGGTCACTGTTGA 431
92 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 109
   |||||:||||| ||| |||:||||| |||
432 CCAGCTCACCAGCAAGCACTCAGGAGCTACAAGTGTGGCGTGGGAGTCA 481
109 suThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 125
   ||| ||| |||: |||: |||: |||
482 AC...GGCCGTGGCGTGGAGCTTCGGTGTCAACGCTGCTGTCAGCCAGAG 528
126 TyrGluPro 128
   |||||
529 CCAGAGCCT 537

seq_name: /cgn_6/ptodata/1/ina/6_COMB.seq:US-08-434-000A-1
seq_documentation_block:
; Sequence 1, Application US/08434000A
; Patent No. 6046037
; GENERAL INFORMATION:
; APPLICANT: ANDREW C. HIATT, JULIAN
; APPLICANT: K.-C. MA, THOMAS LEHNER
; TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
; PROTEINS IN PLANTS AND THEIR USES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/434,000A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below: 1
; APPLICATION NUMBER: 08/367,395
; FILING DATE: 12/30/94
; ATTORNEY/AGENT INFORMATION:
; NAME: Guise, Jeffrey W.
; REGISTRATION NUMBER: 34,613
; REFERENCE/DOCKET NUMBER: 212/127
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 552-8400
; TELEFAX: (619) 552-0159
; TELEX: 67-3510
; SEQUENCE LISTING
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3517 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; TOPOLOGY: DESCRIPTION: Rabbit polyimmunoglobulin receptor
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 124....2445
; US-08-434-000A-1

```

alignment\_scores: Quality: 166.50 Length: 103

```

Ratio: 2.602                                     Gaps: 4
Percent Similarity: 62.136      Percent Identity: 42.718

alignment_block:
US-09-135-238B-2 x US-08-434-000A-1      ..

Align seg 1/1 to: US-08-434-000A-1 from: 1 to: 3517

30  GlyGlySerValThrIleLysCysProLeuPro.....GluMe 42
   ||| |||||:::|||| ||| |||
238  GCGACTCGGTGTCATCAAGCTACTACCCACAACTCCGTCACCCG 287

42  thisValarIleThrLeuCysArgGluMetalaglyserGlyThrCysG 59
   ||| |||::: ||||| ||| |||
288  GCACAGCGCGGAAGTCTCGTGCCGGAA...GAGGAGAGCGCCGCTGC 334

59  1YThrValValserThrThrAsnPhelIeLysAlaGluThrLysGlyArg 75
   |||||::: ||| |||::: :: |||||::: |||||
335  TGACGCTGCCTCG...ACCGGTACACGCTCCAGGAATACTCGGGAGA 381

76  ValThrLeuLysGlnTyProArgLysAsnLeuPheLeuValGluValTh 92
   ||| ||| |||||::: ||||| ||| |||
382  GCGCAAGCTCACGACTCTCCCGATAAAGGGGAGTTGTGTGACTGTGA 431

92  rGlnLeuThrGluSerAspserGlyValTyAlaCysGlyAlaGlyMetA 109
   ||||| ||| ||||| ||||| ||| |||
432  CCNACTCACCAGACGACTCAGGAGCTACAAGTGTGGCGTGGGAGTCA 481

109  snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 125
   || || ||||| ||| |||||::: ||| |||:::
482  AC...GGCCGTGGCTGGACTTGGGTGTCAAGCTGTGGTCAGCCAGAAG 528

126  TyrGluPro 128
   |||||
529  CCAGAGCGCT 537

seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-935-450-10

seq_documentation_block:
; Sequence 10, Application US/08935450
; Patent NO. 5977311
; GENERAL INFORMATION:
; APPLICANT: Nandabalan, Krishnan
; APPLICANT: Yang, Meijia
; APPLICANT: Schulz, Vincent
; TITLE OF INVENTION: 53BP2 COMPLEXES
; FILE REFERENCE: 7934-054
; CURRENT APPLICATION NUMBER: US/08/935,450
; CURRENT FILING DATE: 1997-09-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 10
; LENGTH: 914
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-08-935-450-10

```

```

alignment_scores:
  Quality: 142.00      Length: 247
  Ratio: 1.420        Gaps: 14
  Percent Similarity: 40.486  Percent Identity: 26.721

alignment_block:
  US-09-135-238B-2 x US-08-935-450-10  ..

Align seg 1/1 to: US-08-935-450-10 from: 1 to: 914

133  GlnProMetProGluThrProLybTrpPheHisLeu.....144
232  CAGCCCTGTCGGAGACCCCCCAGGGCCAGCACTCAATAGAACAGCA 281

145  ....ProTyrLeu.....PheGlnMetProAlaTyrAlaIserSers 157

```

```

282 ACATCCCTGGCTCAAGCGCCAAATACAGCAGCCGCCACCTCAGCAGCTAC 331
||||:|||||
157 eLySerPheValThrArgValThrThr.....ProAlaGlnArg..... 169
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
332 AGCCCTTCCACAGCGGATTACAGCGCAGCCACCTACAACCGGGAGGT 381
||||:|||||
170 .....GlyLysValProPro.ValHisHisSerSerP 180
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
382 TACAGCCAGGGTTACAGAGCCCGCCCTCCACCTCCACACCACTGCG 431
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
180 roThrThrGln.....IleThrHisArgProArgValSerArg 192
||||:|||||
432 CTACAACATTATGGGAGCTACGCGGCTTACACCGCGCCCTATACCCCG 481
||||:|||||
193 AlaSerSerValAlaGlyAspLysProArgThrPheLeuProSerThrTh 209
||||:|||||
482 CGCCACGCC.....CCACCGCACAGCACTACCTCAGGCCAACTATAAC 528
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
209 rAlaSerLysIleSerAlaLeuGluGlyLeuLys.....ProG 223
||||:|||||
529 TATCAGCAGTATGCCAGCAGTGAACCACTACTATCAGAACCGGCCAG 578
||||:|||||
223 lInThrProSerTyrAsnHisHisThrArgLeuHisArgGlnArgAlaLeu 239
||||:|||||
579 TGGCGCATACTACGGGAATAC..... 601
||||:|||||
240 AspTyrGlySerGlnSerGlyArgGluGlnGlnGlyPheHisIleLeuII 256
||||:|||||
602 GACTACGGGAGCTACTCTCGGGAACACACAGGGTGGCACAGTACACAGTA 651
||||:|||||
256 eProThrIleLeuGlyLeuPheLeuLeuLeuLeuGlyLeuValVal 273
||||:|||||
652 GCACGTG..... 658
||||:|||||
273 ySerGAlaValGluArgArgLysAlaLeuSerArgArgAlaArgLeu 289
||||:|||||
658 ..... 658
||||:|||||
290 AlaValArgMetArgAlaLeuGluSerSerGlnArgProArgGlySerPr 306
||||:|||||
659 .....TGACCCAGAGGCTCCCG 675
||||:|||||
306 oArgProArgSerGlnAsnAsnIleTySerAlaCys..... 318
||||:|||||
676 GAGGCCCTCGCGGTTCTCCACAC...AGCGCTCGCTCGGCCCTCTCT 722
||||:|||||
319 .....ProArgAlaArgGlyAlaAspAlaGlyThrGlyGluAla 333
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
723 CTGCCCCCGCAGATCCCGTGGTGCT.....GGGATGGG 757
||||:|||||
334 ProValProGly.....ProGlyAlaProLeuProPro 344
||||:|||||
758 GTATCCAGGCGTCTCCCTCCAGCCCACTGCTCC 796
||||:|||||

seq_name: /cgnr2.6/ptodata/1/ina/5D_COMB.seq:US-08-824-405-5

seq_documentation_block:
; Sequence 5, Application US/08824405
; Patent NO. 5962246
; GENERAL INFORMATION:
; APPLICANT: Ladher, Robert D.
; APPLICANT: Lynch, Frank
; APPLICANT: Caradonna, Salvatore J.
; TITLE OF INVENTION: dmpase, Its Isoforms, and
; TITLE OF INVENTION: Diagnostic and Other Uses
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert, Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA

```



REFERENCE/DOCKET NUMBER: ALK88-15AAAZ

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 861-6240

TELEFAX: (617) 861-9540

INFORMATION FOR SEQ ID NO: 27:

SEQUENCE CHARACTERISTICS:

LENGTH: 10785 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)

IMMEDIATE SOURCE:

CLONE: PAH4625

FEATURE:

NAME/KEY: misc\_feature

LOCATION: 1..10785

OTHER INFORMATION: /note= "Function = "Expression"

OTHER INFORMATION: Vector Coding Sequence"

US-08-444-644-27

alignment\_scores:

Quality: 135.00 Length: 446

Ratio: 0.763 Gaps: 23

Percent Similarity: 39.686 Percent Identity: 22.646

alignment\_block:

US-09-135-238B-2 x US-08-444-644-27 ..

Align seg 1/1 to: US-08-444-644-27 from: 1 to: 10785

4 TrpLeuTrpProLeuTyrPheLeuProValSerGlyAlaLeuArgIleLe 20

1854 TGGAGCTGGGTAATGCTCTTCCTC...CTGTCAGGAACTCCAGGTGTCG 1900

20 uProGluValLysValGlu.....GlyGluLeu.....GlyG 31

1901 CTCCTAGGTCGAGCTGCACAGCTCTGGACCTGGACTGGTGAAGCCTGGAG 1950

31 LysSerValThrIleLysCysProLeuProGluMetHisValArgIleTyr 47

1951 CTTCAATGAAGATTCTCTGC..... 1970

48 LeuCysArgGluMetAlaGlySerGlyThrCysGlyThrValValSerTh 64

1971 .....AAGGCTCTGTGTTACTCATCTACTGTC.....TACAC 2002

64 rThrAsnPheIleLys..... 69

2003 CATGAACCTGGGTGAAGCAGAGCCATGGAGAACCTTGATGGATTGGAC 2052

70 .....AlaGluTyrIleGly 74

2053 GTATTAACTCCATCATGTGTGTACTGACTACACAGAGGTTCAAGGAC 2102

75 ArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluVa 91

2103 AAGGCCCTTTTAACCTAGACAAGTCATCCAACACACAGCCATCATGGAGCT 2152

91 lThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGly. 107

2153 CTTCACTGATCTGACATCTGAGGACTCTGCAGTCTATTAATCTGTCAGAGGCT 2202

108 .....MetAsnThrAspArg...GlyLysThrGlnLysValThrLeu 120

2203 ACTATTACTATTCTTTGGACTACTGCGGTCGAAGAACCTCACTCACTGACGTC 2252

121 AsnValHisSerGluTyrGluProSer..... 129

2253 TCCTCAGCTAGCACCAAGGCCCATCGGCTCTCCCTGGGCGCCCTGCTC 2302

130 .....TrpGluG 132

2303 CAGGAGCACTCCGAGAGCACAGCGGCCCTGGCTGCCTGGTCAAGGACT 2352

132 luGlnProMetProGluThrProLysTrpPheHisLeuLeuProTyrLeuPhe 148

2353 ACTTCCCGAAGCGGTGACGGTGTGCTGAA....CTCAGGCGCTCTGAC 2398

149 GlnMetProAlaTyrAlaSerSerLysPheValThrArgValThrTh 165

2399 CAGCGGCTGCACACCTTCCAGCTGCTTACAGTCTCCTCAGGACTCTACT 2448

165 rProAlaGlnArgGly.LysValProValHisHisSerSerPro... 180

2449 CCTCAGCAGCGTGTGACCGTCCCTCCAGCAACTTCGCCACCCAGACC 2498

181 ....ThrThrGlnIleThrHisArgProArgValSerArgAlaSerVa 196

2499 TACCTGCAACCTAGATCACAGCCACCAAGGTGGACAAGAC 2548

196 lAlaGlyAspLysProArgThrPheLeuProSerThrThrAlaSerLysI 213

2549 AGTTGGTGAGAGGCCA.....GCTCAGGGAGGGAGGG 2580

213 leSerAlaLeuGluGlyLeuLeuLysProGlnThrProSerTyrAsnHis 229

2581 TGTCTGCT..... 2588

230 HisThrArgLeuHisArgGlnArgAlaLeuAspTyrGlySerGlnSerG 246

2589 .....GGAAGCCAGGCTCA 2602

246 yArgGluGlyGlnGlyPheHisIleLeuIleProThrIleLeuGlyLeuP 263

2603 G.....CCCTCTGCGCTGGAC.... 2618

263 heLeuLeuAlaLeuLeuGlyLeuValLysArgAlaValGluArgArg 279

2618 ..... 2618

280 LysAlaLeuSerArgArgAlaArgArgLeuAlaValArgMetArgAlaLe 296

2619 .....GCACCCCGGCTGTGCAGCCC 2638

296 uGluSerSerGln.....ArgProArgGlySerProArgProArgSerG 311

2639 CAGCCAGGCGACCAAGGAGCCCATCTCTCTCTCACCAGGAGGCC. 2687

311 lAsnAsnIleTyrSerAlaCysProArgArgAlaArgGlyAlaAsp... 326

2688 .....TCTCCCGCCCTCATCTCATCTCAGGAGGAGGCTCTT 2723

327 .....AlaAlaGlyThrGly..GluAlaProVal 335

2724 CTGCTTTTCCACAGGCTCCAGGAGGAGGAGGCTGGTGGCTCCCTACC 2773

336 ProGlyPro.....GlyAlaProLeuProProAlaProLe 347

2774 CCAGCCCTTCACACAGGAGGAGGAGGCTGTGGCTCAGACCTGCCAAAG 2823

347 uGlnValSerGluSerProTrpIleHisAlaProSerLeu..... 360

2824 CCATATCCGGGAGGACCT.....GCCCTGAGCTAAGCCGACCC 2864

361 ..LysThrSerCysGluTyrValSerLeuTyrHis 371

2865 AAAGGCCAAACGTCCCTCCCTCAGCTCGGACAC 2899

seq\_name: /cgn2\_6/ptodata/1/ina/6\_COMB.seq:US-08-444-644-32

seq\_documentation\_block:

; Sequence 32, Application US/08444644

; Patent No. 6015555

; GENERAL INFORMATION:

; APPLICANT: Friden, Phillip M.







[illegible]

```
4062 CATCAACTGGGTGAGCAGAGCCATGGAGAACCTTGAGTGGATTGGAC 4111
70 .....AlaGluTyrLysGly 74
4112 GTAATATCCTCACAATGGTGGTACTGACTACACCAAGATTCAAGGAC 4161
75 ArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluVa 91
4162 AAGGCCCTTAACTGTAGACAAGTCATCCAAACACAGCCTACATGGAGCT 4211
91 lThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGly. 107
4212 CCTCAGTGTACATCTGAGGACTGTCAGTCTATTACTGTGCAAGAGGCT 4261
108 .....MetAsnThrAspArg...GlyLysThrGlnLysValThrLeu 120
4262 ACTATTACTATTCTTTGGACTACTGGGGTCAAGGAACCTCAGTCACCGTC 4311
121 AsnValHisSerGluTyrGluProSer...Tyr.GluGluGlnProMetP 136
4312 TCCTCAGCTAGCACCAAGGCCCATCGTCTTCCCTCCCTGGCACCTCCTC 4361
136 toGluThrProLysTrpPheHisLeuProTyrLeuPheGlnMetProAla 152
4362 CAAGGACACCTCTGGGGGCACAGCGCCCTGG.....GCTGGCC 4399
153 TyrAlaSerSerLysPheValThrArgValThrThrProAlaGln.. 168
4400 TGG.....TCAGGAGTACTTCCCGCAACCGGTTG 4428
169 .....ArgGlyLysValProProValHisHisSerSerProThrThrG 183
4429 ACGGTGTCGTGGAACCTCAGGGCCCTGACCAGCGCGTGCACACCTTCCC 4478
183 InLeThrHisArgProArgValSerArgAlaSerSerValAlaGlyAsp 199
4479 GGCTGCTTACACTCTCAGGACTCTACTCCCTCAGCAGCGTGTGACCG 4528
200 LysProArgThrPheLeuProSerThrThrAlaSerLysIleSerAlaLe 216
4529 TGCCCT.....CCAGCAGCTTGGGCACCCACAGACCTACATCTGC 4566
216 uGluGlyLeuLysProGlnThrProSerTyrAsnHisHisThrArgL 233
4567 ACGTGAATACAGCCACACCAAGGTGGACA..... 4603
233 euHisArgGlnArgAlaLeuAspTyrGlySerGlnSerGlyArgGluGly 249
4604 .....AGAAAGTTGGTGAGAGCCAGCAC.....AGGGAGGA 4636
250 GlnGlyPheHisIleLeuIleProThrIleLeuGlyLeuPheLeuLeuAl 266
4637 .....GGGTGCTGCTGGAAGC 4653
266 aLeuLeuGlyLeuValValLysArgAlaValGluArgLysAlaLeuS 283
4654 AGGCTCAGCGCTCTGCTGGAGCG...ATCCCGGCTATGCAGCCCGCAGT 4700
283 exArgArgAlaArgLeuAlaValArgMetArgAlaLeuGluSerSer 299
4701 CCAGGGCAGCAAGGCAG.....GCCCGCTCTGCC 4729
300 GlnArgProArgGlySerProArgProArgSerGlnAsnAsnIleTyrSe 316
4730 TCTTACCCGAGGACTCTGCCCGCCCTACTCATGCTCAGGGAGGGTCT 4779
316 rAlaCysProArgArgAlaArgGlyAlaAspAlaAlaGlyThrGlyGluA 333
4780 TCTGGCTTTTCCAGGCTCTGGCAGGCACAGCAGCTAGGT.....G 4820
333 laProValProGlyPro.....GlyAlaProLeuProPro 344
4821 CCCTTAACCCAGGCTCTGCACAAAGGGGAGGTGCTGGGCTCAGACCT 4870
```

```
345 AlaProLeuGlnValSerGluSerProTrpPheHisAlaProSerLeu 360
4871 GCCAAGAGCCATATCGGGAGGACCT.....GCCCTGACCTA 4909
seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-157-101A-6

seq_documentation_block:
; Sequence 6, Application US/08157101A
; Patent No. 5808032
; GENERAL INFORMATION:
; APPLICANT: KURIHARA, TATSUYA
; APPLICANT: MATSUKURA, SHIGEKAZU
; APPLICANT: TSURUOKA, NOBUO
; APPLICANT: ARIMA, KENJI
; APPLICANT: NISHIHARA, TATSURO
; TITLE OF INVENTION: ANTI-HBS ANTIBODY GENES AND EXPRESSION
; TITLE OF INVENTION: PLASMIDS THEREFOR
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PILLSBURY, MADISON & SUTRO
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/157,101A
; FILING DATE: 05-APR-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: TITUS, MARLANA K
; REGISTRATION NUMBER: 35843
; REFERENCE/DOCKET NUMBER: 9437/204199
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-861-3711
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUCH
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1576 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-157-101A-6

alignment_scores:
Quality: 125.50 Length: 432
Ratio: 0.668 Gaps: 24
Percent Similarity: 43.519 Percent Identity: 23.843

alignment_block:
US-09-135-238B-2 x US-08-157-101A-6 ..
Align seg 1/1 to: US-08-157-101A-6 from: 1 to: 1576

6 TrpProLeuTyrPheLeuProValSerGlyAlaLeuArgIleLeuProG1 22
|||||:
17 TGGCCTTTTAAAGAGGTGTCAGTGTCCAGGTGCA..... 49
22 uValLysValGluGly.GluLeuGlyClySerValThrIleLysCysPro 38
|||||:
50 .....GCTGGTGGAGTCTGGGGAGGCGTGGTC..... 77
39 LeuProGluMethHisValArgIle..... 46
|||||:

```

```
78 CAGCCTGGGAGGTCCCTGAGACTCTCTGTGCAGCCTCTGGATTACACTT 127
47 .....TyrLeuCysArgGluMetAlaGlySerGlyThrC 58
128 CAGTAGCAATTCATGCACTGGGTCCGCCAGGCTCCAGGCAAGGGTTGG 177
58 ysGlyThrValVal.....SerThrThrAsnPheIleLysAla 70
178 AGTGGGTGGCAGTTATATATATGATGGAATCATAAATCTACGCACAC 227
71 GluTyrLysGlyArgValThrLeuLysGlnTyrProArgLysAsnLeuPh 87
228 TCCGTGAAGGCCGATTCACATTCATTCAGAGACAATTCACAAGACACACT 277
87 eLeuValGluValThrGlnLeuThrGluSerAspSerGlyValTyrAlaC 104
278 GTATCTGGAGGTGAGAGCCTCAAACTGAGGACACGGGTGCTATTACT 327
104 ysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLysVal..ThrLe 120
328 GT.....ATAAGAGATCAAACTTACGGAGTCCACAGA 359
120 uAsnValHisSerGluTyrGluProSerTrpGluGluGlnProMetPro. 136
360 TTTGACTCTCTGGGGCCAGGGAACCC...TGCTACCGTCTCTCAGCCCTC 406
137 .....GluThrProLysTyrPhe.....HisLeuPro 145
407 CACCAAGGCCCATCGGTCTCCCTCCCTGGCACCTCTCCAGAGACCACT 456
146 TyrLeuPheGlnMetProAlaTyrAlaSerSerLysPheValThrAr 162
457 CTGGGGGCACAGCGCCCTGCTGCTGCTCA.....AG 491
162 gValThrThrProAlaGln.....ArgGlyLysValProProValH 176
492 GACTACTTCCCGAACCGGTGACGGTGTGCTGGAACTCAGCGCCCTGGC 541
176 IsHisSerSerProThrThrGlnIleThrHisArgProArgValSerArg 192
542 CAGCGCGTGCACACCTTCCCGGTGTCTACAGTCTCAGGACTCTACT 591
193 AlaSerSerValAlaGlyAspLysProArgThrPheLeuProSerThrTh 209
592 CCTCAGCAGCGTGTGACCGTGCCT.....CCAGCAGCTTG 629
209 rAlaSerLysIleSerAlaLeuGluGlyLeuLeuLysProGlnThrProS 226
630 GGCACCCAGACCTACATCTGCAAGTGAATCACAAGCCCAACACCAA 679
226 eTyrAsnHisHis.....ThrArgLeu 233
680 GTGGACAAGAAAGTTGAGCCCAATCTTGACAAACATCACACATGCC 729
234 HisArgGlnArgAlaLeu.....AspTyrGlySerGlnSerGl 246
730 CAGCGTGCACGACCTGAACCTCTGGGGGACCGTCACTCTCTCTCTTC 779
246 yArgGlyGlyGlnGlyPheHisIleLeuIleProThrIleLeuGlyLeu 262
780 CCCCCAAACCCAGGACACCTCATGATCTCCCGGACCCCTTGAGGT... 826
263 PheLeuLeuAlaLeuLeuGly..... 269
827 .....CACATGCGTGGTGGTGAGTGAGCCACGAGACCCCTGAGG 867
270 .....LeuValValLysArg..... 274
868 TCAAGTTCAACTGTGTGAGCGCGGTGGAGTGCATAATGCCAAGACA 917
275 .....AlaValGluArgArgLysAlaLeuSerArgArgAla 286
918 AAGCGCGGGAGGAGCAGTACACAGCACCTACCGGGTGTGTCAGCGTCT 967
```

```
287 ArgArgLeuAla.....ValArgMetArgAl 295
968 CACGCTCTGCACCAAGGACTGGCTGAATGGCAAGGAGTACAAGTGCAAGG 1017
295 aLeuGluSerSerGlnArgProArgGlySerProArgProArgSerGlnA 312
1018 TCTCCAAACAAGC.....CCTCCCAAGCCCCCATCGAGAA 1052
312 snAsnIleTyrSer.....AlaCysProArgArgAlaArgGlyAlaAsp 326
1053 ACCATCTCCAAAGCAAGGCGCAGCCCCGAGA..... 1084
327 AlaAlaGlyThrGlyGluAlaProValProGly.....Pr 338
1085 ACCAGAGGTGTACACCTGCCCCCATCCCGGATGAGCTGACCAAGAACC 1134
338 oGlyAlaProLeuProAlaProLeuGlnValSerGlu 351
1135 AGTCAAGCTGACCTGCCTGCTCAAGGCTTCTATCCAG 1174
seq_name: /cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-713-118-1
seq_documentation_block:
; Sequence 1, Application US/08713118
; Patent No. 6040436
; GENERAL INFORMATION:
; APPLICANT: Franco, Rodrigo
; APPLICANT: Sun Chen, Ai Ru
; APPLICANT: Suey, David J.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING HUMAN NEURONAL
; TITLE OF INVENTION: CALCIUM CHANNEL SUBUNITS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173-4799
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08713,118
; FILING DATE: 16-SEP-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mata, Elizabeth W.
; REGISTRATION NUMBER: 38,236
; REFERENCE/DOCKET NUMBER: ACC96-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7266 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 92..7102
; US-08-713-118-1
```

```
alignment_scores:
  Quality: 125.50      Length: 408
  Ratio: 0.721        Gaps: 24
  Percent Similarity: 42.647      Percent Identity: 23.775
```

```
alignment_block:
US-09-135-238B-2 x US-08-713-118-1 ..
Align seg 1/1 to: US-08-713-118-1 from: 1 to: 7266
14 SerGlyAlaLeuArgIleLeuProGluValLysValGluGlyGluLeu1 30
:::||||: ||| |||||:
5710 GCAGGCTCCTGGAGGCTCTCCAGAT.....GG 5738
30 yGlySerValThrIleLysCysProLeuProGluMetHisValArgIle1 47
|||: |||: ||| ||| |||
5739 GTCCGTGTCCTGTT.....CCACCTCTGAGGCCAC..... 5772
47 yLeuCysA:GgIuMetAlaGlySerGlyThrCysGlyThrValValSer 63
||||: ||||: ||||: ||||:
5773 .....CCTGGAGCAGACAGCCGCTGTCTCCGAGG 5805
64 ThrThrAsnPhe.....IleLysAlaGluTyrIy 73
::: ||||: |||: ||||: ||||:
5806 ACCCGGGTTCCTTCGACAGAGTTCACCTCCCTCAGCAATGGCG 5855
73 sGlyArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValG 90
|||: ||||: ||||: ||||: ||||: ||||:
5856 GGGCCATACAAACCAAGAGAGTGGCATCAAGAGTCTGTCTCTGGGGC 5905
90 Lu.....ValThrGlnLeuThrGluSerAspSerGlyVal 101
::: ||||: ||||: ||||: ||||:
5906 ACTCAAGACCAGCATGCCCATCAGGCCAGGCCACCCCTGGAGCG 5955
102 Tyr.....AlaCysGlyAlaGlyMetAsnThrAspArgG1 113
::: ||||: ||||: ||||: ||||:
5956 TGCCCACTCCACAGAGATCCCTGTGGGGCGTCAGGAGCACTGGCTGTG 6005
113 yLysThrGlnLysValThrLeuAsnValHisSerGluTyrGluProSer1 130
|||: ||||: ||||: ||||: ||||: ||||:
6006 AGTTTCAGATGCA.....GAGCATAACCGGAGGGGCCCTGAT 6043
130 rp.GluGluGlnProMetProGluThrProLysTrpPheHisLeuProTy 146
||| ||| ||||| |||||
6044 GGGGAGCCCCAGCTGGCTGGTGAGAGCCAG.....GGTCG 6078
146 rLeuPheGlnMetProAlaTyrAlaSerSerLysPheValThrArgV 163
::: ||||: ||||: ||||: ||||: ||||: ||||:
6079 AGCGCCTCCATGCCCGCTTTCGGCGGAGACTCAGCCGCTCAGAGATG 6128
163 alThrThrProAlaGlnArgGlyLysValProProValHisSerSer 179
::: ||||: ||||: ||||: ||||: ||||: ||||:
6129 CCAGC.....CCCATGAGCGCTCCATC 6151
180 ProThrThrGlnIleThrHisArgProArgValSerArgAlaSerSerVa 196
||| ||| ||||| ||||| ||||| |||||
6152 TCCAGC.....CTGGCCCGAGCGCCCGCTGGGACTCATCTTTGCAGCAC 6195
196 lAlaGlyAspLysProArgThrPheLeuProSerThrThrAlaSerLysI 213
::: ||||: ||||: ||||: ||||: ||||: ||||:
6196 CACCCCGGACCGCCCA.....CCCCCTAGCCAGCGCTGC.... 6229
213 leSerAlaLeuGluGlyLeuLeuLysProGlnThrProSerTyrAsnHis 229
||| ||| ||||| ||||| ||||| |||||
6230 .....TCGCACCCACCAC 6241
230 HisThrArgLeuHisArg.....GlnArgAlaLeuAspTy 241
||| ||| ||||| ||||| ||||| |||||
6242 CACCACCGTCCACCCCGCAGGAGCAGAGCAGAGTCCCTGGAGAA 6291
241 rGlySerGlnSerGlyArgGlu..... 248
||| ||| ||||: ||||: ||||: ||||:
6292 GGGGCCCAAGCTGTCTCGATATGATGGCGGACCAAGCAGTGTGTGG 6341
249 .....GlyGlnGlyPheHisIleLeuIleProThrIle 259
||| ||| ||||: ||||: ||||: ||||:
6342 GGGCGGGGCTGCCCCCGGAGAGGG.....CCTACA... 6373
```

```
260 LeuGlyLeuPheLeuLeuAlaLeuLeuGlyLeuValValLysArgAlaVa 276
6374 .....GGCTGCGCGGGAACG 6390
276 lGluArgGlyLysAlaLeuSerArgAlaArgAlaValArgM 293
||||| ||||| ||||| ||||| |||||
6391 AGAGCCCGGAGAGCGGAGCGGTCCAGGAGCGG.....AGGC 6431
293 etArgAlaLeuGluSerSerGlnArgProArg.GlySerPro..... 306
::: |||: ||||: ||||: ||||: ||||: ||||:
6432 AGCCCTCATCTCTCTCGGAGAACGACGCTTCTACTCTCGACCGC 6481
307 .....ArgProArgSerGlnAsnAsnIleTyrSerAlaCysP 319
||||| ||||| ||||| ||||| |||||
6482 TTTGGGGCGGTGAGCCCCCGAAGCCCAAGCCCTCCCTCAGCAGCCACC 6531
319 roArgArg.....AlaArgGlyAlaAspAlaAlaGlyThrGlyGlu 332
||||| ||||| ||||| ||||| |||||
6532 AAGCTCGCCACACAGCTGCCAGGAGCGGACCC.....ACCCACAGG 6575
333 AlaProValProGlyProGlyAlaPro.....Le 342
||| ||||| ||||| ||||| |||||
6576 GCAGTGTTCCTGAATGGGAGCCCTTGTCTCAACATCTGTGTAGC 6625
342 uProProAlaProLeuGlnValSerGluSerProTrp.....LeuH 356
||||| ||||| ||||| ||||| |||||
6626 ACCCGCGCGGTGGGCGGAGGAGCTCCCCAGACGCCCCCTGACTCC 6675
356 lAlaProSerLeuLysThr 362
||||| ||||| |||||
6676 CCGCCCCAGCATCACCTACA 6695
```

\_\_\_\_\_

•





```

Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 390.
Location/Qualifiers
1. .404
/organism="Homo sapiens"
/db_xref="GDB:6044084"
/db_xref="taxon:9606"
/clone="IMAGE:813174"
/clone_lib="Scorers_NHMPU_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="BHI08"
/note="organ: mixed (see below); Vector: pT73D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not 1;
Site_2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NBHM, pregnant uterus
NBHPU, and fetal heart NBH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."
BASE COUNT 71 a 153 c 114 g 66 t
ORIGIN

alignment_scores:
Quality: 551.00 Length: 111
Ratio: 5.009 Gaps: 0
Percent Similarity: 99.099 Percent Identity: 98.198

alignment_block:
US-09-135-238B-2 x AA456707 .

Align seg 1/1 to: AA456707 from: 1 to: 404

280 LysAlaLeuSerArgAlaArgAlaArgLeuAlaValArgMetArgAlaLeu 296
:::|||||
1 GAAGCCCTCTCCAGCGGGCCCGCGACTGGCGGTGAGGATGCGGCGCT 50

296 uGluSerSerGlnArgProArgGlySerProArgProArgSerGlnAsnA 313
|||
51 GGAGAGCTCCACAGAGGCCCGCGGGTGC CGGA. CCGCGCTCCCAAAACA 99

313 snlleTySerAlaCyProArgArgAlaArgGlyAlaAspAlaGly 329
|||||
100 ACATCTACAGCCCTGCCCGCGCGCGCTCGTGSAGCGAGCTGCAGGC 149

330 ThrGlyGluAlaProValProGlyProGlyAlaProLeuProProAlaPr 346
|||
150 ACAGGGGAGGCGCCGCTGCCGCGCCCGCGAGCGCGCTTGCCTCCCGCCC 199

346 oLeuGlnValSerGluSerProTrpLeuHisAlaProSerLeuLysThrS 363
|||||
200 GCTGAGGTGCTGAAATCTCCCTGGCTCCATGCCCTCTCTGAAGACCA 249

363 eCySGluTyValSerLeuTyHisGlnProAlaAlaMetMetGluAsp 379
|||||
250 GCTGTGAATACGTAGGCTCTACACAGCGCTGCCGCTCATGTGGAGGC 299

380 SerAspSerAspAspTyrlleAsnValProAla 390
|||||
300 AGTGATTACAGTACTACATCAATGTTCTCTGCC 332

seq_name: gb_est42:AW402953
seq_documentation_block:

```



```

LOCUS       AW402953             312 bp      mRNA           EST           16-FEB-2000
DEFINITION   DI-HF-BKO-abd-c-06-0-UI.r1 NIH_MGC_36 Homo sapiens cDNA clone
IMAGE:3055859 5', mRNA sequence.
ACCESSION   AW402953
VERSION     AW402953.1   GI: 6921743
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
             Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 312)
AUTHORS     NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     On May 18, 1998 this sequence version replaced gi:3138550.
             Contact: Robert Strausberg, Ph.D.
             Tel.: (301) 496-1550
             Email: Robert.Strausberg@nih.gov
             Eco RI site shown at the beginning of the sequence.
             Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
             cDNA Library Preparation: M.B. Soares Lab
             CDNA Library Arrayed by: M.B. Soares Lab
             DNA Sequencing by: M.B. Soares Lab
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             www.bio.llnl.gov/brp/image/image.html
             Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1..312
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3055859"
                     /clone_lib="NIH_MGC_36"
                     /tissue_type="lymph"
                     /cell_type="germinal center B cells"
                     /cell_line="MGC85"
                     /lab_host="DH10B (LT1)"
                     /note="Vector: pT7n3-Pac; Site.1: NotI; Site.2: Eco RI;
                     Constructed from size fractionated cytoplasmic mRNA
                     (0.5-1.5kb). Directionally cloned. Cells provided by Louis
                     M. Staudt, Ph.D. Library preparation by Maria de Fatima
                     Bonaldo, Ph.D. and M. Bento Soares, Ph.D."
     BASE COUNT      97 a      74 c      87 g      54 t
     ORIGIN
1  GUMethisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyTh 57
10 GAATGTCATGTGAGGATATATCTGTGCGGAGATGGCTGGATCTGGAAC 59
57 TCysGlyThrValValSerThrThrAsnPheIleLeuValGluTyrLysG 74
60 ATGTGGTACCGTGGTATCCACCACTTCATCAAGGACGAATACAAAG 109
74 lYArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGlu 90
110 GCCGAGTTACTCTGAACATATACCCACCACTTCATCAAGGACGAATACAAAG 159
91 ValThrGlnLeuThrGluSerSerGlyValTyrAlaCysGlyAlaG1 107
160 GTTACACAGCTGACAGAAAGTACACGCGAGTCTATGCTCGGAGCGGG 209
107 yMetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnValHis 124

```

```

210 CATGACACAGACCGGGGAAGACCCAGAAAGTCCACCTGATGTCCACA 259
124 erGluTyrGluProSerTrpGluGlnProMetProGluThrProLys 140
260 GTGAATACGAGCCATCATGGGAAGAGCAGCAATGCTGAGACTCCAAAA 309
141 Trp 141
111
310 TGG 312
seq_name: gb_est45:AW655210
seq_documentation_block:
LOCUS       AW655210             518 bp      mRNA           EST           05-APR-2000
DEFINITION   105863 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   AW655210
VERSION     AW655210.1   GI: 7421036
KEYWORDS    EST.
SOURCE      Bos taurus.
ORGANISM    Bos taurus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
             Bovidae; Bovinae; Bos.
REFERENCE   1 (bases 1 to 518)
AUTHORS     Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
             Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A. and
             Keeler,J.W.
TITLE       Design and use of four pooled tissue normalized cDNA libraries for
             EST discovery in cattle
JOURNAL     Unpublished (2000)
COMMENT     On Mar 8, 1999 this sequence version replaced gi:4388119.
             Contact: Smith TrpL
             USDA, ARS, US Meat Animal Research Center
             PO Box 166, Clay Center, NE 68933-0166, USA
             Tel: 402 762 4366
             Fax: 402 762 4390
             Email: smith@email.marc.usda.gov
             Single pass sequencing. Bases called and alt trimmed with phred
             v0.980904 e. Vector identified by cross_match with the -minscore 18
             and minmatch 12 options.
             PCR Primers
             FORWARD: AGGAACAGCTATGACCAT
             BACKWARD: GTTTCCTCCATCAGCAGC
             Plate: 86 row: K column: 10
             Seq primer: ATTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1..518
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 1BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPOT6; Site_1: XbaI; Site_2: XhoI;
                     library made from pooled tissue from lymph node, ovary,
                     fat, hypothalamus, and pituitary."
     BASE COUNT      124 a      133 c      157 g      104 t
     ORIGIN
alignment_scores:
Quality: 522.00      Length: 136
Ratio: 4.424        Gaps: 0
Percent Similarity: 86.765      Percent Identity: 68.382
alignment_block:
US-09-135-238B-2 x AW655210
Align seg 1/1 to: AW655210 from: 1 to: 518
1 MetAspArgTrpLeuTrpProLeuTyrPheLeuProValSerGlyAlaLe 17
110 ATGAGCTGTGGCTTTGGCGCTTTACTTCTTCCAGTAGTGGAGGCC 159

```





3334-3337, 3682-3683, 3798-3803 (IMAGE Clones)  
 1323376-1323911, 1456008-1456775, 1500552-1502855;  
 NCI\_CGAP\_Kids pool 1 L1AM 3338-3342, 3722-3725, 3776-3778  
 (IMAGE Clones) 1323912-1325831, 1471368-1472903,  
 1492104-1493255); NCI\_CGAP\_Lu5 pool 1 L1AM 3575-3582,  
 3851-3854 (IMAGE Clones)  
 1314920-1417991, 1520904-1522439); NCI\_CGAP\_G04 pool 1 L1AM  
 3164-3167, 3716-3720, 3733-3735 (IMAGE Clones)  
 1357096-158631, 1469064-1470983, 147592-1476743;  
 NCI\_CGAP\_P22 pool 1 L1AM 2457-2459, 2758-2759, 3062-3068  
 (IMAGE Clones) 985608-986759, 1101192-1101959,  
 1217928-1220615); NCI\_CGAP\_C010 pool 1 L1AM 2644-2653,  
 2871-2872 (IMAGE Clones)  
 1057416-1061255, 1144584-1145351). (10% of the driver  
 population), plus a pool of 3,840 arrayed clones from  
 NCI\_CGAP\_Sub1 (IMAGE Clones) 2708616-2710535) and  
 NCI\_CGAP\_Sub2 (IMAGE Clones) 2710536-2712455) (10% of  
 the driver population), plus a pool of 11,136 clones from  
 NCI\_CGAP\_Sub3 (IMAGE Clones) 2712456-2723591) (10% of the  
 driver population), plus a pool of 5,472 clones from  
 NCI\_CGAP\_Sub4 (IMAGE Clones) 2723592-2728969) (70% of the  
 driver population). Subtraction was performed as  
 previously described (Bonaldi, Lennon & Soares (1996):  
 Normalization and Subtraction: Two Approaches To  
 Facilitate Gene Discovery. Genome Research 6, 791-806.  
 TAG\_L1B-NCI\_CGAP\_Kids; TAG\_TISSUE-kidney; TAG\_SEQ-ATTC";

BASE COUNT 103 a 107 c 111 g 161 t  
 ORIGIN

## alignment\_scores:

Quality: 444.50 Length: 117  
 Ratio: 4.401 Gaps: 3  
 Percent Similarity: 86.325 Percent Identity: 80.342

## alignment\_block:

US-09-135-238B-2 x AW444691/rev ..

Align seg 1/1 to reverse of: AW444691 from: 1 to: 482

41 GluMetHisValArgIleTyrLeuGlySerGluMetAlaGlySerGlyTh 57  
 481 GAATCATGATGAGATATATCTGTCCGGGAGATGGCTGATCTGGAC 432  
 57 rCysGlyThrValValSerThrThrAsnPhelIleLysAlaGlyLys 74  
 431 ATGTGTACGGTGTATCCACCACCACTTCATCAAGGAGATACAGG 382  
 74 IyargValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGlu 90  
 381 GCGGATTACTCTGAAGCAATACCCAGCAAGATCTGTCTAGTGGAG 332  
 91 ValThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaG 107  
 331 GTAACACAGCTGACAGAAAGTACACAGCGAGTCTATGCTCGGAGCGG 282  
 107 YMetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnValHis 124  
 281 CATGAACACAGACCGGGAGAACCCAGAAAGTCACCTGATCTCCACA 232  
 124 er...GluTyrGluProSerTyrGluGluGlnProMetProGluThrPro 139  
 231 GTGGTAGGTCTCTCCGCTGATTCGAGGCTCAG.....ACTACCCAG 191  
 140 LysTyrPheHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSerSe 156  
 190 AAAATCTTCTTCATGGGAAGATTG...GAACGCCATCTAATCTGAACAG 144  
 156 r 156  
 143 T 143

seq\_name: gb\_est41:AW327067

## seq\_documentation\_block:

LOCUS AW327067 359 bp mRNA EST 27-JAN-2000  
 DEFINITION 20516 MARC 1BOV Bos taurus CDNA 5', mRNA sequence.  
 ACCESSION AW327067  
 VERSION AW327067.1 GI:6762988  
 KEYWORDS EST.  
 SOURCE Bos taurus.  
 ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 Bovidae; Bovinae; Bos.

1 (bases 1 to 359)

Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,  
 Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A. and  
 Keeler,J.W.

Design and use of four pooled tissue normalized cDNA libraries for  
 EST discovery in cattle

Unpublished (2000)

On Jan 6, 2000 this sequence version replaced gi:6676592.

Contact: Smith TPL

USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366

Fax: 402 762 4390

Email: smith@mail.marc.usda.gov

Single pass sequencing. Bases called and trimmed with phred  
 v0.9804.e. Vector identified by cross\_match with the -minscore 20  
 and -minmatch 12 options.

PCR Primers

FORWARD: AGGAACAGCTATGACCAT

BACKWARD: GTTTCCTCAGTCACGACG

Plate: 13 row: K column: 5

Seq primer: ATTAGTGACACTATAG.

Location/Qualifiers

1..359

/organism="Bos taurus"

/db\_xref="taxon:9913"

/clone\_lib="MARC 1BOV"

/tissue\_type="pooled"

/lab\_host="DH10B"

/note="Vector: PCMV SPORT6; Site:1: XbaI; Site:2: XhoI;  
 Library made from pooled tissue from lymph node, ovary,  
 fat, hypothalamus, and pituitary."

## FEATURES

BASE COUNT 90 a 111 c 87 g 71 t

## alignment\_scores:

Quality: 415.00 Length: 119  
 Ratio: 4.192 Gaps: 1  
 Percent Similarity: 83.193 Percent Identity: 64.706

## alignment\_block:

US-09-135-238B-2 x AW327067 ..

Align seg 1/1 to: AW327067 from: 1 to: 359

76 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 92  
 2 GTCACCTGGACAGTGTCCGGACAGGATCTGTCTTGTGTGTGATGAC 51  
 92 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 109  
 52 AGAGCTGACCAAGAAATGACAGCGGATCTATGCTGTGGGTGGCGGGA 101  
 109 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 125  
 102 ACACAGACCGAGGAGAACCCAGACAGATCAGCTGTTCACAGCGTT 151  
 126 TyrGluProSerTyrGluGluGlnProMetProGluThrProLysTyrPh 142  
 152 TACGAGCCATCTCTGGGAAGAAGCGCATGCTGAGCTCCAGCATGTT 201

```

142 e.....HisLeuProTyrLeuPheGlnMetProAlar 153
202 TAATAGATTCTCAAAATGCACTTCCCTCTGTTTCAGATGCTGCAC 251
153 yralaSerSerfysPheValThrArgValThrProAlaGlnArg 169
252 ATGCCAGTTCTTTAAATTCAGTCCAAAGTAACACACACAGCTCAAGG 301
170 GlyLysValProAlaHisHisSerProThrThrGlnIleThrHI 186
302 ATCAGTCTCTCGACACACACAGGCTCCCGACCCCTCAGTCACCA 351
186 sar9Pro 188
352 CCGCCCC 358
seq_name: gb_est42:AW401870

seq_documentation_block:
LOCUS AW401870 410 bp mRNA EST 16-FEB-2000
DEFINITION UI-HF-BKO-aah-g-11-0-UI-r1 NIH_MGC_36 Homo sapiens cDNA clone
IMAGE:3053781 5', mRNA sequence.
ACCESSION AW401870
VERSION AW401870.1 GI:6920556
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 410)
NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT On Jun 22, 1998 this sequence version replaced gi:3246884.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 495-1550
Email: Robert.Strausberg@nih.gov
Eco RI site shown at the beginning of the sequence.
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
CDNA Library Preparation: M.B. Soares Lab
CDNA Library Arrayed by: M.B. Soares Lab
DNA Sequencing by: M.B. Soares Lab
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: M13 Forward.
FEATURES
Location/Qualifiers
1..410
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3053781"
/clone_lib="NIH_MGC_36"
/tissue_type="lymph"
/cell_type="germinal center B cells"
/lab_host="DH10B (LRI)"
/notes="Vector: pT73-Pac; Site 1: NotI; Site 2: Eco RI;
Constructed from size fractionated cytoplasmic mRNA
(0.5-1.5kb). Directionally cloned. Cells provided by Louis
M. Staudt, Ph.D. Library preparation by Maria de Fatima
Bonaldo, Ph.D. and M. Bento Soares, Ph.D."
BASE COUNT 72 a 154 c 94 g 89 t 1 others
ORIGIN

alignment_scores:
Quality: 372.00 Length: 69
Ratio: 5.391 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-135-238B-2 x AW401870

```

```

Align seg 1/1 to: AW401870 from: 1 to: 410
322 AlaArgGlyAlaAspAlaGlyThrGlyGluAlaProValProGlyPr 338
10 GCTCGTGGAGCGGACGCTGCAGGACAGGAGGAGCCCGCTCCGCC 59
338 oGlyAlaProLeuProAlaProLeuGlnValSerGluSerProTrpL 355
60 CGGAGCGCGCTTGCCTCCCGCCCGCTGCAGGTGTCTGAATCTCC 109
355 euHisAlaProSerLeuLysThrSerCysGluTyrValSerLeuTyrHis 371
110 TCCATGCCCATCTCTGAGACACAGCTGTGATACGTGAGCTTAC 159
372 GlnProAlaAlaMetMetGluAspSerAspSerAspTyrIleAsnVa 388
160 CAGCTGCCCGCATGATGGAGACAGTATTCAGATGACTACATCAT 209
388 lProAla 390
210 TCCTGCC 216
seq_name: gb_est3:AA174968

seq_documentation_block:
LOCUS AA174968 460 bp mRNA EST 16-FEB-1997
DEFINITION ms78H03.r1 Soares mouse 3NBMS Mus musculus cDNA clone IMAGE:617717
5', mRNA sequence.
ACCESSION AA174968
VERSION AA174968.1 GI:1756098
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 460)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:378541
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 459.
FEATURES
Location/Qualifiers
1..460
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:617717"
/clone_lib="Soares mouse 3NBMS"
/sex="male"
/tissue_type="Spleen"
/dev_stage="4 weeks"
/lab_host="DH10B"
/notes="Vector: pT73-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(GT) primer [5',
TGTACCATCTGAGTGGAGCGGCGGCTGTCTTTTCTTTTCTTTTCTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I

```

and Eco RI sites of the modified pT73 vector. RNA provided by Dr. Bertrand Jordan. Library went through three rounds of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 108 a 158 c 117 g 77 t

ORIGIN

alignment\_scores:  
Quality: 360.00 Length: 153  
Ratio: 3.130 Gaps: 5  
Percent Similarity: 75.163 Percent Identity: 54.902

alignment\_block:

US-09-135-238B-2 x AA174968 ..

Align seg 1/1 to: AA174968 from: 1 to: 460

```

188 ProArgValSerArgAlaSerSerValAlaGlyAspLysProArgThrPh 204
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7 CCCAGAGTTTACAGAGCATTTCTGTGTACAGTACCAAGTCCCGAGCGCT 56
204 eLeuProSerThrAlaSerLysIleSerAlaLeuGluGlyLeuLeuL 221
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
57 CCTGCCAGCACCACAGCTCAAGACTTCCACTCAGCAGCA...ATCA 103
221 yProGlnThrProSerTyrAsnHisHisThrArgLeuHisArgGlnArg 237
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
104 GGCCTTAGAGCCAGCTACAGCCACACACAGACTTCATGAGCAAGG 153
238 AlaLeuAspTyrGlySerGlnSerGlyArgGluGlyGlnGly..... 251
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
154 ACAGCCACCATGCCACACTATGGGAGAGAAAGACCGAGGCGTTCACAT 203
252 .....PheHisIleLeuIleProThrIleLeuGlyLeuPheL 264
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
204 CCCCATCCAGAGATTTCCATCTCGACCTCTCTGGGCTTCTCT 253
264 eLeuAlaLeuGlyLeuValValLysArgAlaValGluArgArgLys 280
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
254 TGTGTCTTTTGGGACTGGTGTAAAGAGCCATTCAGAGGAGGA 303
281 AlaLeuSerArgArgAlaArgArgLeuAlaValArgMetArgAlaLeuG 297
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
304 GCTCTCCAGAGCTGGCGCCGACTACGATGAGGAGCGGCGCGGG 353
297 uSerSerGln.Arg.ProArgGlySerPro.....ArgProAr 309
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
354 GGTCTCCGCGCGTTCCTCCACACAGCGCGCGATGCTCGCAGACGCGG 403
309 gSerGlnAsnAsnIleTyrSerAlaCysProArgArgAlaArgGlyAla 326
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
404 CTCGAGAGAACAGCTCTACAGCGCTGCCCCGCGCGCGGACGACG 453
326 spLa 327
|||||
454 ACAGC 458

```

seq\_name: gb\_est4:AA289667

seq\_documentation\_block:

LOCUS AA289667 417 bp mRNA EST 14-APR-1997  
DEFINITION vb20c07.r1 Soares mouse 3NDMS Mus musculus cDNA clone IMAGE:749484  
5', mRNA sequence.  
ACCESSION AA289667  
VERSION AA289667.1 GI:1937343  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 417)  
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubouque,T.,

Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lannon,G., Soares,B., Wilson,R. and Waterston,R.  
The WashU-HHMI Mouse EST Project  
Unpublished (1996)  
Contact: Marra M/Mouse EST Project  
WashU-HHMI Mouse EST Project  
Washington University School of MedicineP  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LML; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
MG1:458468  
Seq primer: -28ml3 rev2 ET from Amersham  
High quality sequence stop: 405.  
Location/Qualifiers  
1. .417  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:749484"  
/clone\_lib="Soares mouse 3NDMS"  
/sex="male"  
/tissue\_type="Spleen"  
/dev\_stage="4 weeks"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGCGCTGTCTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. RNA provided by Dr. Bertrand Jordan. Library went through three rounds of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."

FEATURES  
source

BASE COUNT 109 a 140 c 91 g 77 t  
ORIGIN

alignment\_scores:  
Quality: 349.00 Length: 138  
Ratio: 3.231 Gaps: 2  
Percent Similarity: 78.261 Percent Identity: 53.623

alignment\_block:  
US-09-135-238B-2 x AA289667 ..

Align seg 1/1 to: AA289667 from: 1 to: 417

```

162 ArgValThrThrProAlaGlnArgGlyLysValProValHisHis 178
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4 AAAGTTACACAGCCCAACTCCAAAGACTGAGGCCCTCCGGTTCACGACC 53
178 rSerProThrThrGlnIleThrHisArgProArgValSerArgAlaSerS 195
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
54 CTCAGCATCCTCTCAGTAAGCCCAACATCCAGAGTTTACAGAGCATTT 103
195 erValAlaGlyAspLysProArgThrPheLeuProSerThrThrAlaSer 211
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
104 CTGTGTGAGTACCAAGTCCCGAGCGCTCTGCCAGCAACACAGCCCTCA 153
212 LysIleSerAlaLeuGluGlyLeuLysProGlnThrProSerTyrAs 228
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
154 AAGACTTCCACTCAGCAAGCA...ATCAGGCCCTCAGAGCCAGCTACAG 200
228 nhisHisThrArgLeuHisArgGlnArgAlaLeuAspTyrGlySerGlnS 245
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
201 CCACACACACAGACTCTATGAGCAAAAGACAGCCACCATGCGCCCACT 250

```





www-bio.llnl.gov/bbrp/image/image.html  
 Insert Length: 983 Std Error: 0.00  
 Seq primer: -40ml3 fwd. RT from Amersham  
 High quality sequence stop: 473.

## FEATURES

source  
 Location/Qualifiers  
 1..485

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1353883"  
 /clone\_lib="NCI\_CGAP\_GCB1"  
 /tissue\_type="germinal center B cell"  
 /lab\_host="DH10B"  
 /note="vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
 was prepared from human tonsillar cells enriched for  
 germinal center B cells by flow sorting (CD20+, IgD-),  
 provided by Dr. Louis M. Staudt (NCI), Dr. David Allman  
 (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was  
 primed with a Not I - oligo(dT) primer  
 [5'-TGTTACCAATCTGAAGTGGAGGCGCGCTCATTTTTTTTTTTT-3'  
 ]. Double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pT73 vector. Library  
 went through one round of normalization, and was  
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 109 a 84 c 137 g 155 t  
 ORIGIN

alignment\_scores:  
 Quality: 305.00 Length: 59  
 Ratio: 5.169 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-135-238B-2 x AA830944/rev ..

Align seg 1/1 to reverse of: AA830944 from: 1 to: 485

179 SerProThrGlnIleThrHisArgProArgValSerArgAlaSerSe 195  
 |||||  
 485 TCCCCACCCACCAATCACCACCGCCCTCGAGTGTCCAGAGCATCTC 436  
 195 rValAlaGlyAspLysProArgThrPheLeuProSerThrThrAlaSerL 212  
 |||||  
 435 AGTAGCAGGTGACAGCCCGGAACCTCTCTGCCATCTCCTACAGCCTCAA 386  
 212 ystleSerAlaLeuGluGlyLeuLysProGlnThrProSerTyrAsn 228  
 |||||  
 385 AAATCTCAGCTCTGGAGGGGCTCTCAAGCCGCCAGAGCCGCTACAA 336  
 229 HisHisThrArgLeuHisArgGlnArg 237  
 |||||  
 335 CACACACACGAGCTGCACAGGCAGAGG 309

seq\_name: gb\_est8:AA491123

seq\_documentation\_block:  
 LOCUS AA491123 179 bp mRNA EST 18-AUG-1997  
 DEFINITION aa46904.r1 NCI\_CGAP\_GCB1 Homo sapiens cDNA clone IMAGE:824022 5',  
 mRNA sequence.

ACCESSION AA491123  
 VERSION AA491123.1 GI:2220296

KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 179)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index

JOURNAL Unpublished (1997)

## COMMENT

Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,  
 Ph.D., Gerald Marti, M.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
 Bonaldo, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -28ml3 rev1 RT from Amersham

High quality sequence stop: 116.

## FEATURES

Location/Qualifiers

source

1..179

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:824022"

/clone\_lib="NCI\_CGAP\_GCB1"

/tissue\_type="germinal center B cell"

/lab\_host="DH10B"

/note="vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
 was prepared from human tonsillar cells enriched for  
 germinal center B cells by flow sorting (CD20+, IgD-),  
 provided by Dr. Louis M. Staudt (NCI), Dr. David Allman  
 (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was  
 primed with a Not I - oligo(dT) primer  
 [5'-TGTTACCAATCTGAAGTGGAGGCGCGCTCATTTTTTTTTTTT-3'  
 ]. Double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pT73 vector. Library  
 went through one round of normalization, and was  
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 46 a 69 c 34 g 30 t  
 ORIGIN

alignment\_scores:  
 Quality: 299.00 Length: 59  
 Ratio: 5.068 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-135-238B-2 x AA491123 ..

Align seg 1/1 to: AA491123 from: 1 to: 179

163 ValThrThrProAlaGlnArgGlyLysValProValHisHisSerSe 179  
 |||||  
 3 GTTACCACACCAAGTCAAGGGCAAGGTCTCCAGTTCACCACTCTC 52  
 179 rProThrThrGlnIleThrHisArgProArgValSerArgAlaSerSe 196  
 |||||  
 53 CCCCACCCACCAATCACCACCGACCTCGAGTGTCCAGAGCATCTTCAG 102  
 196 alaAlaGlyAspLysProArgThrPheLeuProSerThrThrAlaSerLys 212  
 |||||  
 103 TAGCAGGTGACAAAGCCGCAAGCTTCTCCATCCACTACAGCTCAAAA 152  
 213 IleSerAlaLeuGluGlyLeuLys 221  
 |||||  
 153 ATCTCAGCTCTGGAGGGGCTGCTCAAG 179

seq\_name: gb\_est40:AW291041

seq\_documentation\_block:

LOCUS AW291041 561 bp mRNA EST 16-JAN-2000

DEFINITION UI-H-B12-ag1-h-06-0-UI.s1 NCI\_CGAP\_Sub4 Homo sapiens cDNA clone  
 IMAGE:2724707 3', mRNA sequence.

ACCESSION AW291041

VERSION AW291041.1 GI:6697677



KEYWORDS

SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
TITLE 1 (bases 1 to 561)  
JOURNAL NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
COMMENT National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On May 18, 1998 this sequence version replaced gi:3137211.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
The sequence contained an oligo-dn track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:  
NCI-CGAP clone distribution information can be found through the  
I.M.A.G.E. Consortium/LNL at:  
[www.bio.lnl.gov/bbrp/image/image.html](http://www.bio.lnl.gov/bbrp/image/image.html)  
Seq primer: M13 Forward  
POLYA=Yes.

FEATURES

source  
Location/Qualifiers  
1..561  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2724707"  
/clone\_lib="NCI-CGAP\_Sub4"  
/lab\_host="DH10B (Life technologies)"  
/note="Vector: pTT3D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not 1; Site\_2: Eco RI; The  
NCI-CGAP\_Sub4 library is a subtracted library derived from  
the NCI-CGAP\_Sub2 library which is a subtracted library  
derived from the NCI-CGAP\_Sub1 library, which is a  
subtracted library derived from BI. Bi constitutes a  
mixture of 21 normalized or subtracted NCI-CGAP  
libraries: NCI-CGAP\_Co4, NCI-CGAP\_Pr22, NCI-CGAP\_Pr28,  
NCI-CGAP\_Co10, NCI-CGAP\_Co16, NCI-CGAP\_Kid5,  
NCI-CGAP\_Kid12, NCI-CGAP\_Kid3, NCI-CGAP\_Kid11,  
NCI-CGAP\_Lym2, NCI-CGAP\_Br2, NCI-CGAP\_Co8, NCI-CGAP\_CLL1,  
NCI-CGAP\_Lym2, NCI-CGAP\_Brn23, NCI-CGAP\_Lu5,  
NCI-CGAP\_Lu24, NCI-CGAP\_Lu19, NCI-CGAP\_GG4, NCI-CGAP\_GC6,  
NCI-CGAP\_Brn25. These 21 libraries were pooled and a  
single-stranded DNA preparation of the resulting mixture  
was used as a tracer in a subtractive hybridization with  
a driver whose composition is detailed below:  
NCI-CGAP\_Kid3 pool 1 : LLAM 3334-3337, 3682-3683,  
3798-3803 (IMAGE Clonoids 1322376-1323911,  
1456008-1456775, 1500552-1502855) NCI-CGAP\_Kid5 pool 1 :  
LLAM 3338-3342, 3722-3725, 3776-3778 (IMAGE Clonoids  
1323912-1325831, 1471368-1472903, 1492104-1493255)  
NCI-CGAP\_Lu5 pool 1 : LLAM 3575-3582, 3851-3854 (IMAGE  
Clonoids 1414920-1417991, 1520904-1522439) NCI-CGAP\_GC4  
pool 1 : LLAM 3164-3167, 3716-3720, 3733-3735 (IMAGE  
Clonoids 1257096-1258631, 1469064-1470983,  
1475592-1476743) NCI-CGAP\_Pr22 pool 1 : LLAM 2457-2459,  
2758-2759, 3062-3088 (IMAGE clonoids 985608-986759,  
1101192-1101959, 1217928-1220615) NCI-CGAP\_Co10 pool 1 :  
LLAM 2644-2653, 2871-2872 (IMAGE Clonoids 1057416-1061255,  
1144584-1145351) Subtraction was performed as previously  
described [Bonaldo, Lennon & Soares (1996): Normalization  
and Subtraction: Two Approaches To Facilitate Gene  
Discovery. Genome Research 6, 791-806.];  
TAG\_LIB=NCI-CGAP\_Co10; TAG\_TISSUE=colon; TAG\_SEQ=AAACG"  
121 a 125 c 120 g 194 t 1 others

alignment\_scores:

Quality: 292.50 Length: 91  
Ratio: 3.953 Gaps: 3  
Percent Similarity: 81.319 Percent Identity: 72.527

alignment\_block:

US-09-135-238B-2 x AW291041/rev ..  
Align seg 1/1 to reverse of: AW291041 from: 1 to: 561  
67 PheIleLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProAr 83  
561 TTCATCAGGCAGATACAGGCCCGAGTTACTCTGAAGCAATACCCACG 512  
83 gLysAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerG 100  
511 CAAGAATCTGTTCTAGTGGAGGTAACACAGCTGACAGAAAGTACACGC 462  
100 lyValTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGln 116  
461 GAGTCTATGCTCGGAGCGGGCATGAACACAGACCGGNGAAGACCCAG 412  
117 LysValThrLeuAsnValHisSer...GluTyrGluProSerTrpGluG1 132  
411 AAAGTCACCTGAATGTCCACAGTGGTAGGTTCTCCGCTGATTGGAGGC 362  
132 uGlnProMetProGluThrProLysTrpPheHisLeuProTyrLeuPheG 149  
361 TCAG.....ACTACCCAGAAAAATGTTTCACCTGGGAAGATTG...G 324  
149 InMetProAlaTyrAlaSerSer 156  
323 AACAGCCATCTAACTGTAAACAGT 301

-----

OM of: US-09-135-238B-2 to: Issued\_Patents\_NA:\* out\_format : pfs  
Date: Sep 12, 2000 6:53 AM  
About: Results were produced by the GenCore software, version 4.5.  
Copyright (c) 1993-2000 Compugen Ltd.

## Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-O=/cgn2\_1/USPFO.spool/US09135238/runat\_24082000\_091324\_17895/app\_query.fasta\_1.1282  
-DB=Issued\_Patents\_NA -QFMT=fastap -SUFFIX=p2n.rni  
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000  
-LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000  
-XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000  
-YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1  
-MATRIX=blosum62 -TRANS=human40 cdi -LIST=45 -DOCALIGN=200  
-THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTENT=pfs -NORM=ext -MINLEN=0 -MAXLEN=100000  
-USER=US09135238 @cgn1\_1\_68 -NCP0=6 -ICPU=3 -LONGLOG -NO\_XLPXY  
-WAIT -THREADS=1

## Search information block:

Query: US-09-135-238B-2  
Query length: 390  
Database: Issued\_Patents\_NA.\*  
Database Sequences: 243080  
Database length: 68777915  
Search time (sec): 93.130000

## score.list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-434-000A-7	1	198.50	262.73	3.8e-07	3095
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-434-000A-9	1	185.50	243.58	4.5e-06	3269
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-434-000A-3	1	182.00	239.68	7.4e-06	2919
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-434-000A-5	1	182.00	237.20	1.0e-05	3630
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-642-406A-21	1	166.50	221.76	7.3e-05	1875
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-642-406A-6	1	166.50	221.59	0.0002	3517
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-434-000A-1	1	166.50	214.59	0.0002	3517
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-935-450-10	1	142.00	193.65	0.0027	914
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-824-405-5	1	136.50	178.54	0.0187	1683
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-444-644-27	1	135.00	155.12	0.3776	10785
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-444-644-32	1	131.00	147.85	0.9588	12127
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-378-939-9	1	130.00	169.36	0.0608	1617
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-444-644-18	1	129.50	146.21	1.18	11528
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-157-101A-6	1	125.50	162.98	0.1377	1576
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-713-118-1	1	125.50	145.55	1.29	7266
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-286-740-3	1	124.50	145.24	1.34	6557
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US95-09576-3	1	124.50	145.24	1.34	6557
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-576-628A-2	1	123.00	140.67	2.41	8051
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-476-273-2	1	123.00	139.14	2.93	9209
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-478-967A-3	1	123.00	139.14	2.93	9209
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-819-866-2	1	123.00	130.88	8.45	18986
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-09-023-715-2	1	123.00	130.88	8.45	18986
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-480-036-1	1	122.50	154.29	0.4200	2287
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-461-968A-1	1	122.50	154.29	0.4200	2287
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-462-571-1	1	122.50	154.29	0.4200	2287
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US96-10043-8	1	122.50	154.29	0.4200	2287
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-916-917-1	1	122.50	146.09	1.20	4692
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-972-631-1	1	122.50	146.09	1.20	4692
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-972-629-1	1	122.50	146.09	1.20	4692
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-972-630-1	1	122.50	146.09	1.20	4692
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-672-211-1	1	122.50	146.09	1.20	4692
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-09-225-170-1	1	122.50	146.09	1.20	4692
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-455-543A-7	1	121.50	139.47	2.81	7362
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-193-078B-7	1	121.50	139.47	2.81	7362
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-223-305C-7	1	121.50	139.47	2.81	7362
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-149-097D-7	1	121.50	139.47	2.81	7362
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-310-370-1	1	121.00	159.52	0.2147	1190
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-955-937A-1	1	119.50	149.56	0.7705	2345
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-444-644-41	1	119.50	132.09	7.24	10844

/cgn2\_6/ptodata/1/ina/5B\_COMB.seq:US-08-764-233A-1 - 118.50 113.31 80.51 49377  
/cgn2\_6/ptodata/1/ina/5C\_COMB.seq:US-08-455-543A-8 + 118.00 134.57 5.26 7175  
/cgn2\_6/ptodata/1/ina/5C\_COMB.seq:US-08-193-078B-8 + 118.00 134.57 5.26 7175  
/cgn2\_6/ptodata/1/ina/5C\_COMB.seq:US-08-223-305C-8 + 118.00 134.57 5.26 7175  
/cgn2\_6/ptodata/1/ina/5C\_COMB.seq:US-08-149-097D-8 + 118.00 134.57 5.26 7175

seq\_name: /cgn2\_6/ptodata/1/ina/6\_COMB.seq:US-08-434-000A-7

seq\_documentation\_block:  
: Sequence 7, Application US/08434000A  
: Patent No. 6046037  
: GENERAL INFORMATION:  
: APPLICANT: ANDREW C. HIATT, JULIAN  
: APPLICANT: K.-C. MA, THOMAS LEHNER  
: TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION  
: TITLE OF INVENTION: PROTEINS IN PLANTS AND THEIR USES  
: NUMBER OF SEQUENCES: 19  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Lyon & Lyon  
: STREET: 633 West Fifth Street  
: CITY: Suite 4700  
: CITY: Los Angeles  
: STATE: California  
: COUNTRY: U.S.A.  
: ZIP: 90071

## COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/434,000A

## FILING DATE:

CLASSIFICATION: 435

## PRIOR APPLICATION DATA:

PRIOR APPLICATION DATA: including application

## PRIOR APPLICATION DATA:

PRIOR APPLICATION DATA: described below: 1

## FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Guise, Jeffrey W.

REGISTRATION NUMBER: 34,613

REFERENCE/DOCKET NUMBER: 212/127

## TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 552-8400

TELEFAX: (619) 552-0159

## TELEX:

TELEX: 67-3510

## INFORMATION FOR SEQ ID NO: 7:

## SEQUENCE CHARACTERISTICS:

LENGTH: 3095 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

TOPOLOGY: DESCRIPTION: Mouse Polyimmunoglobulin Receptor

## FEATURE:

NAME/KEY: Coding Sequence

LOCATION: 85.....2400

US-08-434-000A-7

## alignment\_scores:

Quality: 198.00 Length: 168  
Ratio: 1.980 Gaps: 5  
Percent Similarity: 59.524 Percent Identity: 31.548

## alignment\_block:

US-09-135-238B-2 x US-08-434-000A-7 ..

Align seg 1/1 to: US-08-434-000A-7 from: 1 to: 3095

10 PheLeuProValSerGlyAlaLeuArgIleLeuProGluValValG1 26

```

121 TTTTCAGGGGCTCCCAAAAAACCCATATTTGGTCCCGAGAGGTGAG 170
26 uGlyGluLeuGlyGlySerValThrIleLysCysProLeuProGlu... 41
171 TAGTATAGAGGGGACCTCTGTTCCATCAGCTGCTACTACCCAGACACCT 220
42 .....MetHisValArgIleTyrLeuCysArgGluMetAlaGlySer 55
221 CTGTACAGCGCACACCGGAATCTGTCGCGCAGACAGAGGCC...AGC 267
56 GlyThrCysGlyThrValSerThrAsnPhelIleLysAlaGluTyr 72
268 GCGATGTCACAGGCTCATCTCTTCAATGGCTACCTCTCCAAAGAGTA 317
72 rLysGlyArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuV 89
318 TTCAGCAGAGCCAACTCATCACTCCCGAGAGAACACACATTGTGA 367
89 aGluValThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGly 105
368 TPACATTCACACCTCACCAGAGACGACCTGGAGCTACAGTGTGC 417
106 AlaGlyMetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnVa 122
418 CTGGGT...ACCAGTAACCGAGGCGCTCTCTCGATGTCAGCCTGGAGST 464
122 lHisSerGluTyrGluProSerTrpGluGlnProMetProGluThrP 139
465 C.....AGCCAGGTTCCTCGAGTGC 484
139 rGlyTrpPheHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSer 155
485 CGAGTGACACCCAGCTC.....TACACAAAG 510
156 SerSerLysPheValThrArgValThrProAlaGlnArgGlyLysVa 172
511 GACATAGCAGAAATGTGACCAATTGAATGCCCTTCAAAAGGGAGATGT 560
172 lPro 173
561 TCOC 564

seq_name: /cgn2_6/ptodata/L/ina/6_COMB.seq:us-08-434-000A-9
seq_documentation_block:
; Sequence 9, Application US/08434000A
; Patent No. 6046037
; GENERAL INFORMATION:
; APPLICANT: ANDREW C. HIATT, JULIAN
; APPLICANT: K.-C. MA, THOMAS LEHNER
; TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
; TITLE OF INVENTION: PROTEINS IN PLANTS AND THEIR USES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LYON & LYON
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/434,000A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

```

```

; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below: 1
; FILING DATE: 12/30/94
; ATTORNEY/AGENT INFORMATION:
; NAME: Guise, Jeffrey W.
; REGISTRATION NUMBER: 34,613
; REFERENCE/DOCKET NUMBER: 212/127
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 552-8400
; TELEFAX: (619) 552-0159
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3269 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 74....2383
; US-08-434-000A-9

alignment_scores:
Quality: 185.50 Length: 164
Ratio: 1.855 Gaps: 6
Percent Similarity: 60.976 Percent Identity: 30.488

alignment_block:
US-09-135-238B-2 x US-08-434-000A-9 ..
Align seg 1/1 to: US-08-434-000A-9 from: 1 to: 3269
30 GlyGlySerValThrIleLysCysProLeuProGlu.....Me 42
170 GGTAACTCGGTCCTCCATCAGCTGCTACTACCCAGACACCTCTGTCAACGG 219
42 thisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 59
220 GCACACCGGAAATACTGTCGCCGACAGAGGCC...AAGGCTACTCGG 266
59 lYThrValValSerThrAsnPhelIleLysAlaGluTyrLysGlyArg 75
267 CAACCCCTCATCTCTCAATGGCTACCTCTCGAAGGAGTATTGAGGAGA 316
76 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 92
317 GCCAGCCTCATCACTCCCGAGAGAAATAGTGTGATTGATTAAACATTGC 366
92 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 109
367 ACATCTCACCCAGAGGACACTGGGACCTACAGTGTGCTGGGT...A 413
109 sThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 125
414 CCACCTAACCGAGGCGCTGTTTTCATGATCGAGCTGGAGGTC..... 454
126 TyrGluProSerTrpGluGlnProMetProGluThrProLysTrpPh 142
455 .....AGCCAGGTTCCTGAGTTCCTCCAAATGACAC 483
142 eHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSerSerSerLysP 159
484 CCATGTC.....TACACAAAGG 500
159 heValThrArgValThrThrProAlaGlnArgGlyLysValProProVal 175
501 ACATAGCAGCACTGTGACCAATGCTGCTTCAAAAGAGGGAATGCT 550
176 HisHisSer_SerProThrThrGlnIleThrHisArgPro 188

```





```

382  GGCAAGCTCACCGACTCCCTGATAAAGGGAGTTTGTCAGCTGTGA 431
      ||| :||||:|||| |||:|: |||:||||| |||
92  rGlnLeuThrGluSerGlyValTyrAlaCysGlyAlaGlyMetA 109
      |||:|||||:|||||:||||| ||| |||:|||||:|||||:
432  CCACTCACCAAGACGACTCAGGAGCTCAAGTGTGGCGTGGAGTCA 481
      |||:|||||:|||||:||||| ||| |||:|||||:|||||:
109  snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 125
      || ||||| |||||:|||||: ||| :|||||
482  AC...GGCGGTGGCGCTGGACTTCGGTGTCAACGTGCTGGTCAGCCAGAAG 528
      ||||| |||||:|||||: ||| :|||||
126  TyrGluPro 128
      |||||
529  CCAGAGCCT 537

seq_name: /cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-434-000A-1
seq_documentation_block:
; Sequence 1, Application US/08434000A
; Patent No. 6046037
; GENERAL INFORMATION:
; APPLICANT: ANDREW C. HIATT, JULIAN
; APPLICANT: K.-C. MA, THOMAS LEHNER
; TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
; TITLE OF INVENTION: PROTEINS IN PLANTS AND THEIR USES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/434,000A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below:
; APPLICATION NUMBER: 09/4327,405

```

ATTORNEY/AGENT INFORMATION:  
NAME: Guise, Jeffrey W.  
REGISTRATION NUMBER: 34,613  
REFERENCE/DOCKET NUMBER: 212/127  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 552-8400  
TELEFAX: (619) 552-0159  
TELEX: 67-3510

```
;      TELEX: SEQUENCE LISTING
;      INFORMATION FOR SEQ ID NO. 1.
```

```
;
; SEQUENCE CHARACTERISTICS:
;     LENGTH: 3517 base pairs
;     TYPE: nucleic acid
```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; TOPOLOGY: DESCRIPTION: Rabbit polyclonal immunoglobulin receptor

```

```

; NAME/KEY: Coding Sequence
; LOCATION: 124...2445
FEATURE:

```

US-08-434-000A-1

alignment_scores:	Quality:	Length:
	166.50	103

Ratio: 2.602 Gaps: 4  
Percent Similarity: 62.136 Percent Identity: 42.718

## alignment\_block:

US-09-135-238B-2 x US-08-434-000A-1 ..

Align seg 1/1 to: US-08-434-000A-1 from: 1 to: 3517

```
30 GlyGlySerValThrIleLysCysProLeuPro.....GluMe 42
   |||||:|||||:|||||
238 GCGCACTCGGTGTCATCATCTACTACCAACAACCTCCCTCACC 287
   ||| |||: ||| |||||
42 ThisValArgIleThrLeuCysArgGluMetAlaGlySerGlyThrCysG 59
   ||| |||: ||| |||||
288 GCACAGCGCGAAGTCTCGTCCCGGAA...GAGGAGAGCGCGCTCGG 334
   |||:||||| |||:|||||
59 IyThrValValSerThrAsnPheIleLysAlaGluThrLysGlyArg 75
   |||:||||| |||:|||||
335 TGACGCTTGCTCG...ACGCGCTACACGTCACAGGAGTCTCCGGGAGA 381
   ||| |||: ||| |||||
76 ValThrLeuLysGlnTyProArgLysAsnLeuPheLeuValGluValTh 92
   ||| |||: ||| |||||
382 GCGAAGCTCACCGACTCCCTGATAAAGGGAGTTGTGTGAGTGTGA 431
   ||||| |||: ||| |||||
92 rGlnLeuThrGluSerAspSerGlyValTyAlaCysGlyAlaGlyMeta 109
   ||||| |||: ||| |||||
432 CCAACTCACCAAGACGACTCAGGAGCTACAAGTGTGGCGTGGAGTCA 481
   ||||| |||: ||| |||||
109 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 125
   || ||||| |||: ||| |||||
482 AC...GCCCTGCGCTGGACTTCGGTGTCAACGTGTGTGTCAGCCAGAAG 528
   || |||||
126 TyrGluPro 128
   |||||
529 CCAGAGCCT 537
```

seq\_name: /cgn2\_5/ptodata/1/ina/5D\_COMB.seq:US-08-935-450-10

## seq\_documentation\_block:

```
; Sequence 10, Application US/08935450
; Patent No. 5977311
; GENERAL INFORMATION:
; APPLICANT: Nandabalan, Krishnan
; APPLICANT: Yang, Meifia
; APPLICANT: Schuiz, Vincent
; TITLE OF INVENTION: 53BP2 COMPLEXES
; FILE REFERENCE: 7934-054
; CURRENT APPLICATION NUMBER: US/08/935,450
; CURRENT FILING DATE: 1997-09-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 10
; LENGTH: 914
; TYPE: DNA
; ORGANISM: Homo sapiens
US-08-935-450-10
```

alignment\_scores:  
Quality: 142.00 Length: 247  
Ratio: 1.420 Gaps: 14  
Percent Similarity: 40.486 Percent Identity: 26.721

## alignment\_block:

US-09-135-238B-2 x US-08-935-450-10 ..

Align seg 1/1 to: US-08-935-450-10 from: 1 to: 914

```
133 GlnProMetProGluThrProLysTrpPheHisLeu..... 144
   ||||| ||| |||||
232 CAGCCCTGCTCGGAACCCCGCCAGGCGCAGCACTACAATAAGAACAGCA 281
   |||||
145 ....ProTyrLeu.....PheGlnMetProAlaTyAlaSerSers 157
```

```
282 ACATCCCTGGCTCAAGCGCAATACCAAGCCGCCACCGCTCAGCAGTAC 331
   |||:||||| :||| |||
157 erLysPheValThrArgValThr.....ProAlaGlnArg..... 169
   : ||| :||| |||||: |||
332 AGCCCTCCACAGCCGAGTTACGCCAGCCACCTACACACGAGGAGGT 381
   ||| :||| |||||: |||
170 .....GlyLysValProPro.ValHisHisSerSerP 180
   ||| |||: ||| |||||
382 TACAGCCAGGTTTACACAGGCCCGCTCCACCTCCACCCACCTGTC 431
   ||| |||: ||| |||||
180 rofThrGln.....IleThrHisArgProArgValSerArg 192
   ||| |||: ||| |||||
432 CTACACTATGGGAGCTACGGCGTTACACCGCGCCCTATACCCAC 481
   ||| |||: ||| |||||
193 AlaSerSerValAlaGlyAspLysProArgThrPheLeuProSerThrTh 209
   : ||| : |||
482 CGCCACCCC...CCACCGCACAGACTACCTCAGCCCACTATAACCCAG 528
   ||| |||: ||| |||||
209 rAlaSerLysIleSerAlaLeuGluGlyLeuLys.....ProG 223
   : |||: ||| |||: ||| |||||
529 TATCAGCAGTATGCCAGCAGTGGACCACTACTATCAGAACCGGCCAG 578
   ||| |||: ||| |||||
223 lnThrProSerTyrAsnHisThrArgLeuHisArgGlnArgAlaLeu 239
   : ||| : |||
579 TGGCGCATACTACGGGAATAC..... 601
   ||| |||: ||| |||||
240 AspTyrGlySerGlnSerGlyArgGluGlyGlnGlyPheHisIleLeuIl 256
   ||| |||: ||| |||||
602 GATACGGAGGACTACTCCGGGAACACACAGGTTGGCACAGTACACAGTA 651
   ||| |||: ||| |||||
256 eProThrIleLeuGlyLeuPheLeuAlaLeuLeuGlyLeuValValL 273
   ||| |||: ||| |||||
652 GCCAGTG..... 658
   ||| |||: ||| |||||
273 ysArgAlaValGluArgArgLysAlaLeuSerArgArgAlaArgLeu 289
   ||| |||: ||| |||||
658 ..... 658
   ||| |||: ||| |||||
290 AlaValArgMetArgAlaLeuGluSerSerGlnArgProArgGlySerPr 306
   ||| |||: ||| |||||
659 .....TGACCCAGAGGCTCCCG 675
   ||| |||: ||| |||||
306 oArgProArgSerGlnAsnIleTyrSerAlaCys..... 318
   ||| |||: ||| |||||
676 GAGGCCCTGCGCGCTTCTCCACC...AGGCCCTGCTCGGCCCTCCT 722
   ||| |||: ||| |||||
319 .....ProArgArgAlaArgGlyAlaAspAlaAlaGlyThrGlyGluAla 333
   ||| |||: ||| |||||
723 CTGCCCGCCAGATCCCGTGTCT.....GGGATGGG 757
   ||| |||: ||| |||||
334 ProValProGly.....ProGlyAlaProLeuProPro 344
   ||| |||: ||| |||||
758 GTATCCAGGCGCTCCTCCCTCCAGCCCACTGCTCCC 796
   ||| |||: ||| |||||
```

seq\_name: /cgn2\_6/ptodata/1/ina/5D\_COMB.seq:US-08-824-405-5

## seq\_documentation\_block:

```
; Sequence 5, Application US/08824405
; Patent No. 5962246
; GENERAL INFORMATION:
; APPLICANT: Ladner, Robert D.
; APPLICANT: Lynch, Frank
; APPLICANT: Caradonna, Salvatore J.
; TITLE OF INVENTION: dUTPase, Its Isoforms, and
; TITLE OF INVENTION: Diagnostic and Other Uses
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert, Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
```



```
; ZIP: 19103-2793
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/824,405
; FILING DATE: 26-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 350163-101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609/520-3214
; TELEFAX: 609/520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1683 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-824-405-5

alignment_scores:
  Quality: 136.50      Length: 265
  Ratio: 1.034         Gaps: 15
  Percent similarity: 49.811  Percent Identity: 25.660

alignment_block:
US-09-135-238B-2 x US-08-824-405-5/rev ..
Align seg 1/1 to reverse of: US-08-824-405-5 from: 1 to: 1683

162 ArgValThrThrProAlaGln.....ArgGlyLysValProPro...Va 175
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1524 CGGAGCGCGCGCCCTTCCAGCCCTCTCCGCGCGGAGTCCCGCTCACCT 1475
175 lHisHisSerSerProThrThrGlnlleThrHisArgProArgValSerA 192
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1474 GTACAGGTCTGTAGCCCGCGCGCGCG.....GAGCCCGCGGTGGGG 1431
192 rgAlaSerSerValAlaGlyAspLysProArgThrPheLeuProSerThr 208
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1430 CCGTGGCGTCTCGAGAGCGCGGCAAGCGGAGCTGCATGCCGCCACC 1381
209 ThrAlaSerLys.....lleSerAlaLe 216
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1380 TCCGAGCGCGCGCGCGCTTACTGGTGAATGGCGGTGTCTTTCAGA 1331
216 uGlyLeuLeuLysProGlnThrProSerTyrAsnHisHis..... 230
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1330 GCAGGGCATGCAGCAGAGAGAGCGGAGGAGACCCGCGGGGAAGA 1281
231 .....ThrArgLeuHisArgGlnArgAlaLeuAspTyrGly 242
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1280 GGAGGCCACGATGAGGTCTCGCGCGGAGCGCGCCAAACCGAAATT 1231
243 SerGln..... 244
|||||:
1230 TCCGCCAGCCCGCCGCCACCAGCGCCCTTCCCGGAGCAGTCCACAG 1181
245 SerClyArgGluGlyGlnGlyPheHislleLeuIleProThrIleLeuG 261
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1180 TTCTGGTGAGAGGGAGCGGGGATTTCAC.....ATTTGG 1146
261 lyLeu.....PheLeuLeuAlaLeuGlyLeu..... 270
```

```
||||| 1145 GTTTGAAGCCCTTCTCGAAAGCAGAGGTTTGGGATGGGAAACTACTGGGTC 1096
||||| 271 .....ValValLysArgAlaValGluArgArgLysAl 281
||||| 1095 CCAGTCAACTTTCACATTCAGCGCGTGGGCCAACATTGGAAGGGGAGCT 1046
||||| 281 aLeuSerArgArgAlaArgArgLeuAlaValArgMetArgAlaLeuGln 298
||||| 1045 TCTAGGGCGTGC.....CTCACACACTAGAAAGCCGCCCT 1008
||||| 298 erSerGlnArgProArgGlySerProArg.....ProArgSerGln 311
||||| 1007 CGCCAGGGGCCCTTAATGGACCACCCGCCGCCGCCGCCGCCGCCGAG 958
||||| 312 AsnAsn.....lleTyrSerAlaCysProArgArgAlaArgG 324
||||| 957 AACCAACCGCTCGCCACCCCTACTTCCACACAGCCTCAAGCGCGGTGATTC 908
||||| 324 yAlaAspAlaAlaGlyThrGlyGluAlaPro...ValProGlyProGlyA 340
||||| 907 CTTCCAGAGCGCGGAGCCCTCCCGCCCTTCTTACCGCGCCCGCGG 858
||||| 340 laProLeuProProAla.....ProLeuGln..... 348
||||| 857 CGGGGCTTCCCGCCCTTAGGAAGCTCGCCCTTCCAGCAGCGGCCCG 808
||||| 349 .....ValSerGluSerProThrPheHisAlaProSerLeu 360
||||| 807 ACTGTACTGTGCTCGCGGCGAGCCTTGGCTCAGCGCGCGCGCTG 763

seq_name: /cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-444-644-27
seq_documentation_block:
; Sequence 27, Application US/08444644
; Patent No. 6015555
; GENERAL INFORMATION:
; APPLICANT: Friden, Phillip M.
; TITLE OF INVENTION: TRANSFERRIN RECEPTOR SPECIFIC
; TITLE OF INVENTION: ANTIBODY-NEUROPHARMACEUTICAL OR DIAGNOSTIC AGENT
; TITLE OF INVENTION: CONJUGATES
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,644
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/232,246
; FILING DATE: 07-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/800,458
; FILING DATE: 26-NOV-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US90/05077
; FILING DATE: 07-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/404,089
; FILING DATE: 07-SEP-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Wagner, Richard W.
; REGISTRATION NUMBER: 34,480
```

REFERENCE/DOCKET NUMBER: ALK88-15AAAZ

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 861-6240

TELEFAX: (617) 861-9540

INFORMATION FOR SEQ ID NO: 27:

SEQUENCE CHARACTERISTICS:

LENGTH: 10785 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)

IMMEDIATE SOURCE:

CLONE: PAH4625

FEATURE:

NAME/KEY: misc.feature

LOCATION: 1..10785

OTHER INFORMATION: /note= "Function = "Expression

OTHER INFORMATION: Vector Coding Sequence"

US-08-444-644-27

alignment\_scores:

Quality: 135.00 Length: 446

Ratio: 0.763 Gaps: 23

Percent Similarity: 39.686 Percent Identity: 22.646

alignment\_block:

US-09-135-238B-2 x US-08-444-644-27 ..

Align seg 1/1 to: US-08-444-644-27 from: 1 to: 10785

4 TrpLeuProMetPheLeuProValSerGlyAlaLeuArgIleLeu 20 \*

1854 TGAGCTGGGTAAAGCTCTTCCTC...CTGCAGGAAGTGCAGGTGCGG 1900

20 uProGluValLysValGlu.....GlyGluLeu.....GlyG 31

1901 CTCTGAGGTCCAGCTGCAACAGCTCTGGACCTGGAACCTGGTGAAGCCTGGAG 1950

31 LySerValThrIleLysCysProLeuProGluMetHisValArgIleLys 47

1951 CTTCATGAGATTCCTGTC.....GlyGluLeu.....GlyG 1970

48 LeuCysArgGluMetAlaGlySerGlyThrCysGlyThrValValSerTh 64

1971 .....AAGGCTTCTGTTACTCAATTCACCTGGC.....TACAC 2002

64 rThrAsnPheIleLys.....GlyGluLeu.....GlyG 69

2003 CATGAACCTGGGTGAAGCAGAGCATGGAGAACCTTGAGTGGATTGGAC 2052

70 .....AlaGluThrLysGly 74

2053 GTATTATCTCACAATGGTGGTACTACACAGAGAGTTCAGGAC 2102

75 ArgValThrLeuLysGlnTyrProArgLysAsnLeuPheValGluVal 91

2103 AAGGCCCTTAACTGTAGCAAGTCAATCCACACAGCCCTACATGAGCT 2152

91 lThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGly 107

2153 CCTCAGTCTGACATCTGAGGACTCTGCAGTCTATTACTGTGCAAGAGCT 2202

108 .....MetAsnThrAspArg.....GlyLysThrGlnLysValThrLeu 120

2203 ACTATTACTATCTTTGGACTACTGGGTCAAGGAACCTCAGTCACCGTC 2252

121 AsnValHisSerGluTyrGluProSer.....GlyGluLeu 129

2253 TCTCAGTACAGCAACAGGCCCATCGGTCTTCCCTCGGCCCTGCTGTC 2302

130 .....TrpGluG 132

2303 CAGGAGCACCCTCGAGAGCAGAGCGGCCCTGGGCTGCCGTCAAGGACT 2352

132 luGlnProMetProGluThrProLysTrpPheHisLeuProTyrIleuPhe 148

2353 ACTTCCCGAACCGGTGAGGTGTGCTGGAA...CTCAGGCGCTCTGAC 2398

149 GlnMetProAlaTyrAlaSerSerSerLysPheValThrArgValThrTh 165

2399 CAGCGCGTGCACACCTTCCAGCTGTCCAGTGTCTACAGTCTCAGGACTTACT 2448

165 rProAlaGlnArgGly.LysValProValHisHisSerSerPro... 180

2449 CCTCAGCAGCGTGTGACCGTCCAGCAACTTCGGCAACCCAGACC 2498

181 ....ThrThrGlnIleThrHisArgProArgValSerArgAlaSerSerVa 196

2499 TACACCTGCACGTAGTACACAGCCCAACACCAAGGTGGACAAGAC 2548

196 lAlaGlyAspLysProArgThrPheLeuProSerThrThrAlaSerLysI 213

2549 AGTTGGTGAGAGGCCA.....GCTCAGGAGGAGGAGG 2580

213 lSerAlaLeuGluGlyLeuLysProGlnThrProSerTyrAsnHis 229

2581 TGTCTGCT..... 2588

230 HistThrArgLeuHisArgGlnArgAlaLeuAspTyrGlySerGlnSerGl 246

2589 .....GGAAGCCAGGCTCA 2602

246 yArgGluGlyGlnGlyPheHisIleLeuIleProThrIleLeuGlyLeuP 263

2603 G.....CCTCTCTGCTGCCTGCAC.... 2618

263 hLeuLeuAlaLeuLeuGlyLeuValValLysArgAlaValGluArgArg 279

2618 ..... 2618

280 LysAlaLeuSerArgArgAlaArgAlaValArgMetArgAlaLe 296

2619 .....GCACCCCGGCTGTGCAGCC 2638

296 uGluSerSerGln.....ArgProArgGlySerProArgProArgSerG 311

2639 CAGCCAGGCGCAGCAGGAGCCCATCTCTCTCTCCTCACCAGGAGGCC. 2687

311 lAsnAsnIleTyrSerAlaCysProArgArgAlaArgGlyAlaAsp... 326

2688 .....TCTGCCGCCCTCACTCATGCTCAGGAGAGGGTCTT 2723

327 .....AlaAlaGlyThrGly..GluAlaProVal 335

2724 CTGGCTTTTCCAGCAGGCTCCAGGAGCAGGCTGGGTGGCCCTTACC 2773

336 ProGlyPro.....GlyAlaProLeuProProAlaProLe 347

2774 CAGGCGCTTCCACACAGAGGCGAGGTGTGCTGCTCAGACCTGCCAAG 2823

347 uGlnValSerGluSerProTyrLeuHisAlaProSerLeu..... 360

2824 CCATATCCGGGAGGACCT.....GCCCTGACCTTAGCCGACCC 2864

361 ..LysThrSerGysGluTyrValSerLeuTyrHis 371

2865 AAAGGCCAAACTGTCCACTCCTCCTCAGCTGGGACAC 2899

seq\_name: /cgn2\_6/ptodata/1/ina/6\_US\_COMB.seq:US-08-444-644-32

seq\_documentation\_block:

; Sequence 32, Application US/08444644

; Patent No. 6015555

; GENERAL INFORMATION:

; APPLICANT: Friden, Phillip M.



[illegible]

seq name: /cqn2 6/ptodata/1/ina/5C COMB.seq:US-08-378-939-9

```

seq.documentation_block:
: Sequence 9, Application US/08378939
: Patent No. 5876961
: GENERAL INFORMATION:
: APPLICANT: CROWE, JAMES SCOTT
: APPLICANT: LEWIS, ALAN PETER
: TITLE OF INVENTION: PRODUCTION OF ANTIBODIES
: NUMBER OF SEQUENCES: 46
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: ROTHWELL, FIGG, ERNST & KURZ
: STREET: 555 THIRTEENTH ST. N.W.
: CITY: WASHINGTON
: STATE: D. C.
: COUNTRY: U.S.
: ZIP: 20004
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/378,939
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/952640
: FILING DATE: 01-DEC-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: ERNST, BARBARA G.
: REGISTRATION NUMBER: 30,377
: REFERENCE/DOCKET NUMBER: 1808-118
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 783-6040
: TELEFAX: (202) 783-6031
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1617 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: both
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: HYPOTHEICAL: NO
: ANTI-SENSE: NO

```



```

4062 CATGAACCTGGGTGAAGCAGACGACCATGGAGAACCTTGAGTGGATTGGAC 4111
70 .....AlaGluThrLysGly 74
4112 GTATTAACTCCCAATGGTGGTACTGACTACAAACAGAGAGTTCAAGGAC 4161
75 ArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluVal 91
4162 ARGGCCCTTTAACTGTPAGCAAGTCATCCAAACACAGCCTACATGGAGCT 4211
91 lThrGlnLeuThrGlnSerAspSerGlyValTyrAlaCysGlyAlaGly. 107
4212 CCTCAGCTGTGACATCTGAGGACTCTGCGACTCTATTACTGTGCAAGAGGCT 4261
108 .....MetAsnThrAspArg...GlyLysThrGlnLysValThrLeu 120
4262 ACTATTACTATCTTTGGACTACTGGGTCAAGGAACCTCAGTCACCGTC 4311
121 AsnValHisSerGlnTyrGlnProSer...Tyr.GluGlnGlnProMetP 136
4312 TCCTCAGCTAGCAACCAAGGCCCATCGGTCTTCCCTGGGCACCTCCTC 4361
136 roGlnThrProLysTyrPheHisLeuProTyrLeuPheGlnMetProAla 152
4362 CAAGAGCACCTCTGGGGCACAGCGGCCCTGG.....GCTGCC 4399
153 TyrAlaSerSerLysPheValThrArgValThrThrProAlaGln... 168
4400 TGG.....TCAGGACTACTTCCCGACCGGTG 4428
169 .....ArgGlyLysValProProValHisHisSerSerProThrThrG 183
4429 ACGGTGCTGGAACTCAGCGCCCTGACGCGGTGCGACACTTCCC 4478
183 lntleThrHisArgProArgValSerArgAlaSerSerValAlaGlyAsp 199
4479 GGCTGTCTACAGCCCTCAGGACTCTACTCCTCAGCAGCGTGGTGACCG 4528
200 LysProArgThrPheLeuProSerThrThrAlaSerLysIleSerAlaLe 216
4529 TGCCT.....CCAGCAGCTGGGCCACCCAGACCTACATCTGC 4566
216 uGluGlyLeuLysProGlnThrProSerTyrAsnHisHisThrArgL 233
4567 AGGTGAATCAAGCCCAACCAACCAAGGTGGACA.....AGGGAGGGA 4603
233 euHisArgGlnAlaLeuAspTyrGlySerGlnSerGlyArgGluGly 249
4604 .....AGAAAGTTGGTGAGAGCCAGCAC.....AGGGAGGGA 4636
250 GlnGlyPheHisIleLeuIleProThrIleLeuGlyLeuPheLeuAl 266
4637 .....GGGTGCTGCTGGAGC 4653
266 aLeuLeuGlyLeuValValLysArgAlaValGluArgArgLysAlaLeuS 283
4654 AGGCTCAGCGCTCTGCTGGAGCG...ATCCCGGCTATGACAGCCCGCAGT 4700
283 erArgArgAlaArgAlaValArgMetArgAlaLeuGluSerSer 299
4701 CCAGGCGCAGAGCAG.....GCCCGCTCTGCC 4729
300 GlnArgProGlySerProArgProArgSerGlnAsnAsnIleTyrSe 316
4730 TCTTACCGGAGGCTCTGCCGCCCCCTCAGTCAGGAGAGGCTCT 4779
316 rAlaCysProArgArgAlaArgGlyAlaAspAlaAlaGlyThrGlyGlu 333
4780 TCGGCTTTTCCAGGCTCTGGCAGCAGCAGCAGCTAGT.....G 4820
333 laProValProGlyPro.....GlyAlaProLeuProPro 344
4821 CCCCTAACCCAGGCCCTGACACAAAGGGGCGAGTCTGGGCTCAGACCT 4870

```

```

345 AlaProLeuGlnValSerGluSerProTyrLeuHisAlaProSerLeu 360
4871 GCCAAGACCATATCCGGGAGGACCT.....GCCCTGACCTA 4909
seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-157-101A-6
seq_documentation_block:
; Sequence 6, Application US/08157101A
; Patent NO. 5808032
; GENERAL INFORMATION:
; APPLICANT: KURIHARA, TATSUYA
; APPLICANT: MATSUKURA, SHIGEKAZU
; APPLICANT: TSURUOKA, NOBUO
; APPLICANT: ARIMA, KENJI
; APPLICANT: NISHIHARA, TATSURO
; TITLE OF INVENTION: ANTI-HBS ANTIBODY GENES AND EXPRESSION
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PILLSBURY, MADISON & SUTRO
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/157,101A
; FILING DATE: 05-APR-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: TITUS, MARLANA K
; REGISTRATION NUMBER: 35843
; REFERENCE/DOCKET NUMBER: 9437/204199
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-661-3711
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUCH
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1576 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-157-101A-6
alignment_scores:
Quality: 125.50 Length: 432
Ratio: 0.668 Gaps: 24
Percent Similarity: 43.519 Percent Identity: 23.843
alignment_block:
US-09-135-238b-2 x US-08-157-101A-6 ..
Align seq 1/1 to: US-08-157-101A-6 from: 1 to: 1576
6 TrpProLeuTyrPheLeuProValSerGlyAlaLeuArgIleLeuProGl 22
17 TGGCTTTTAAAGAGGTGCCAGTGTCTAGGTGCA..... 49
22 uValLysValGluGly.GluLeuGlyGlySerValThrIleLysCysPro 38
50 .....GCTGTGGAGTCTGGGGAGGCGTGGTC..... 77
39 LeuProGluMethHisValArgIle..... 46

```

```

287 ArgArgLeuLeuAa.....ValArgMetArgal 295
      :|||: ||| |||:||||:||||:
968 CACCCTCTCTGACCAAGGAGTGGCTGAATGGCAAGGAGTACAAGTGCAGG 1017
      :|||:||||:||||:||||:||||:||||:
295 aleuAluSerSerGlnArgProArgProArgProArgProArgSerGlnA 312
      :||||:||||:||||:||||:||||:||||:
1018 TCTCCAAACAAGC.....CCTCCAGCCCCCATCGAGAA 1052
      :|||:||||:||||:||||:||||:
312 sNaSniletyrSer.....AlaCysProArgArgAlaArgGlyAlaAsp 326
      :|||: ||| |||:||||:||||:
1053 ACCATCTCCAAAGCAAGGCGAGCCCGAGA..... 1084
      :|||:||||:||||:||||:||||:
327 AlaAlaGlyThrGlyGluAlaProValProGly.....Pr 338
      :||||:||||: |||:||||:||||:
1085 ACCACAGGTGTACACCTCCCTCCCGGATGAGTGCACCAAGAAC 1134
      :|||:||||:||||:||||:||||:
338 oglyAlaProLeuProAlaProLeuGlnValSerGlu 351
      :|||: ||| |||:||||:||||:
1135 AGTCAGCTGACCTGCGCTGGTCAAGGCTTCTATCCAG 1174
      :|||: ||| |||:||||:||||:

seq_name: /cgn2_5/ptodata/1/ina/6_COMB.seq:US-08-713-118-1

seq_documentation_block:
; Sequence 1, Application US/08713118
; Patent No. 6040436
; GENERAL INFORMATION:
; APPLICANT: Franco, Rodrigo
; APPLICANT: Sun Chen, Ai Ru
; APPLICANT: Suey, David J.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING HUMAN NEURONAL
; TITLE OF INVENTION: CALCIUM CHANNEL SUBUNITS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173-4799
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/713,118
; FILING DATE: 16-SEP-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mata, Elizabeth W.
; REGISTRATION NUMBER: 38,236
; REFERENCE/DOCKET NUMBER: ACC96-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7266 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 92..7102
US-08-713-118-1

alignment_scores:
Quality: 125.50 Length: 408
Ratio: 0.721 Gaps: 24
Percent Similarity: 42.647 Percent Identity: 23.775

```

```

alignment_block:
US-09-135-238b-2 x US-08-713-118-1  ..
Align seg 1/1 to: US-08-713-118-1 from: 1 to: 7366

14 SerGlyAlaLeuArgIleLeuProGluValLysValGluGlyGluLeuG1 30
:::|||||::: ||| |||||||:::
5710 GCAGGCTCTGGAGCGCTCTCCAGAT.....GG 5738
:::|||||::: ||| |||||||:::

30 yGlySerValThrIleLysCysProLeuProGluMetHisValArgIleT 47
::: ||| :::: ||| ||| |||
5739 GTCTGTGTCTCTGT.....CCACCTCTGAAGGCCAC..... 5772
::: ||| :::: ||| ||| |||

47 yrLeuCysArgGluMetAlaGlySerGlyThrCysGlyThrValValSer 63
|||||::: |||::: |||::: |||:::
5773 .....CCTGGAGCAGACAGCGCGCTGTGCTCCGAGG 5805
::: |||::: |||::: |||::: |||:::

64 ThrThrAsnPhe.....IleLysAlaGluTyr-Ly 73
::: |||::: |||::: |||::: |||:::
5806 AGCCCGGCTTTTCCTTCGACAGAAGAGTTCCACCTCTCCAGCAATGGCG 5855
::: |||::: |||::: |||::: |||:::

73 sGlyArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValG 90
|||||::: |||::: |||::: |||::: |||:::
5856 GGCCATACAAACACAGAGTGGCATCAAGAGTCTGTCTCTCGGGGC 5905
::: |||::: |||::: |||::: |||::: |||:::

90 lu.....ValThrGlnLeuThrGluSerAspSerGlyVal 101
|||||::: |||::: |||::: |||::: |||:::
5906 ACTAAAGGCCAGGATGCACCCCATGAGGCCAGGCCACCCCTGGAGCG 5955
::: |||::: |||::: |||::: |||::: |||:::

102 Tyr.....AlaCysGlyAlaGlyMetAsnThrAspArgL 113
::: |||::: |||::: |||::: |||::: |||:::
5956 TGCCACTCCACAGATCCTGTGGGGGTGTCAGGACACTGGCTGTGG 6005
::: |||::: |||::: |||::: |||::: |||:::

113 YLysThrGlnLysValThrLeuAsnValHisSerGluTyrGluProSerT 130
|||||::: |||::: |||::: |||::: |||:::
6006 ACCTTCAGATGCA.....GAGCATAACCCGGAGGGGCCCTGAT 6043
::: |||::: |||::: |||::: |||::: |||:::

130 rp.GluGluGlnProMetProGluThrProLysTrpPheHisLeuProTy 146
||| ||| ||||||| |||:::
6044 GGGAGCCCCAGCTGGCTGGAGAGCCAG.....GGTCG 6078
::: |||::: |||::: |||::: |||::: |||:::

146 rLeuPheGlnMetProLysThrAlaSerSerLysPheValThrArgV 163
::: |||::: |||::: |||::: |||::: |||:::
6079 AGCGGCCTCCATGCCCGCTTGCGGCCGAGACTCAGCCGCTCAGATG 6128
::: |||::: |||::: |||::: |||::: |||:::

163 alThrThrProAlaGlnArgGlyLysValProProValHisHisSerSer 179
::: |||::: |||::: |||::: |||::: |||:::
6129 CCAGC.....CCCATGAAGCGCTCCATC 6151
::: |||::: |||::: |||::: |||::: |||:::

180 ProThrThrGlnIleThrHisArgProArgValSerArgAlaSerSerVa 196
||| ||| ||||||| |||::: |||:::
6152 TCCAGC.....CTGGCCAGCGCGCCGTGGGACTCATCTTTCAGCAGC 6195
::: |||::: |||::: |||::: |||::: |||:::

196 lAlaGlyAspLysProArgThrPheLeuProSerThrThrAlaSerLysI 213
::: |||::: |||::: |||::: |||::: |||:::
6196 CACCCCGGAGCGGCCA.....CCCCCTAGCCAGGCGTCG.... 6229
::: |||::: |||::: |||::: |||::: |||:::

213 lSerAlaLeuGluGlyLeuLeuLysProGlnThrProSerTyrAsnHis 229
|||||::: |||::: |||::: |||::: |||:::
6230 .....TCGACCCACCAC 6241
::: |||::: |||::: |||::: |||::: |||:::

230 HisThrArgLeuHisArg.....GlnArgAlaLeuAspTy 241
||| ||| ||||||| |||::: |||:::
6242 CACCACCGCTCCACCGCCGAGGACAGGAGCAGAGGTCCCTGGAGAA 6291
::: |||::: |||::: |||::: |||::: |||:::

241 rGlySerGlnSerGlyArgGlu..... 248
||| :::: |||::: |||::: |||::: |||:::
6292 GGGGCCAGCGCTGTGCCGATATGGATGGCGCACCAACAGTGTGTGG 6341
::: |||::: |||::: |||::: |||::: |||:::

249 .....GlyGlnGlyPheHisIleLeuIleProThrIle 259
|||||::: |||::: |||::: |||::: |||:::
6342 GCGCGGGCTGCCCGGGGAGAGGG.....CCTACA... 6373
::: |||::: |||::: |||::: |||::: |||:::

```

```

260 LeuGlyLeuPheLeuAlaLeuLeuGlyLeuValValLysArgAlaVa 276
6374 .....GGCTGCCGCGCGGAACG 6390
::: |||::: |||::: |||::: |||::: |||:::

276 lGluArgArgLysAlaLeuSerArgAlaArgArgLeuAlaValArgM 293
|||||::: |||::: |||::: |||::: |||:::
6391 AGAGCGCGCAGAGCGAGCGCGTCCAGAGCGG.....AGGC 6431
::: |||::: |||::: |||::: |||::: |||:::

293 eArgAlaLeuGluSerSerGlnArgProArg GlySerPro..... 306
::: |||::: |||::: |||::: |||::: |||:::
6432 AGCCCTCATCTCTCTCGGAGAGCAGCGCTTCTACTCTCTGCGACCGC 6481
::: |||::: |||::: |||::: |||::: |||:::

307 .....ArgProArgSerGlnAsnAsnIleTyrSerAlaCysP 319
|||||::: |||::: |||::: |||::: |||:::
6482 TTTGGGGCGCGTGAGCCCGGAGCCCAAGCCCTCTCAGCAGCCACC 6531
::: |||::: |||::: |||::: |||::: |||:::

319 rOArgArg.....AlaArgGlyAlaAspAlaAlaGlyThrGlyGlu 332
|||||::: |||::: |||::: |||::: |||:::
6532 AACGTCCGCCAACAGCTGGCCAGGAGCGGAGCCCC.....ACCCACAGG 6575
::: |||::: |||::: |||::: |||::: |||:::

333 AlaProValProGlyProGlyAlaPro.....Le 342
||| ||||| |||||||
6576 GCAGTGGTTCGTGAATGGAGCCCTTGTCTCAACATCTGTGTCTAGC 6625
::: |||::: |||::: |||::: |||::: |||:::

342 uProProAlaProLeuGlnValSerGluSerProTrp.....LeuH 356
|||||::: |||::: |||::: |||::: |||:::
6626 ACCCCCGCGCGGTGGGCGGAGCGAGCTCCCCAGACGCCCTGACTCC 6675
::: |||::: |||::: |||::: |||::: |||:::

356 lAlaProSerLeuLysThr 362
|||||::: |||::: |||::: |||::: |||:::
6676 CCGCCCCAGCATCACCTACA 6695
::: |||::: |||::: |||::: |||::: |||:::

```



' . . '

Tue Sep 12 08:57:55 2000

us-09-135-238b-2.p2n.rni

Page 15

a



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2000, 10:09:19 ; Search time 3690.46 Seconds  
(without alignments)  
923.671 Million cell updates/sec

Title: US-09-135-238B-1  
Perfect score: 1910  
Sequence: 1 aaagagtaagcagcgtgctc.....ttactctgtccatccttt 1910

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 972840 seqs, 892348106 residues  
Total number of hits satisfying chosen parameters: 1945680

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- GenEmbl:\*
- 1: gb\_bal:\*
  - 2: gb\_ba2:\*
  - 3: gb\_om:\*
  - 4: gb\_ov:\*
  - 5: gb\_pat:\*
  - 6: gb\_ph:\*
  - 7: gb\_pl1:\*
  - 8: gb\_pl2:\*
  - 9: gb\_pri:\*
  - 10: gb\_pr2:\*
  - 11: gb\_pr3:\*
  - 12: gb\_ro:\*
  - 13: gb\_sts:\*
  - 14: gb\_sy:\*
  - 15: gb\_un:\*
  - 16: em\_fun:\*
  - 17: em\_hum1:\*
  - 18: em\_hum2:\*
  - 19: em\_in:\*
  - 20: em\_on:\*
  - 21: em\_or:\*
  - 22: em\_ov:\*
  - 23: em\_pat:\*
  - 24: em\_ph:\*
  - 25: em\_pl:\*
  - 26: em\_ro:\*
  - 27: em\_sts:\*
  - 28: em\_sy:\*
  - 29: em\_un:\*
  - 30: em\_vi:\*
  - 31: gb\_in1:\*
  - 32: gb\_in2:\*
  - 33: gb\_in3:\*
  - 34: gb\_pl3:\*
  - 35: gb\_pr4:\*
  - 36: gb\_pr5:\*
  - 37: gb\_htg1:\*
  - 38: gb\_htg2:\*
  - 39: gb\_htg3:\*
  - 40: gb\_htg4:\*
  - 41: gb\_htg5:\*
  - 42: gb\_htg6:\*
  - 43: gb\_htg7:\*

- 44: gb\_htg8:\*
- 45: gb\_htg9:\*
- 46: gb\_htg10:\*
- 47: gb\_htg11:\*
- 48: gb\_htg12:\*
- 49: gb\_htg13:\*
- 50: gb\_htg14:\*
- 51: gb\_htg15:\*
- 52: gb\_htg16:\*
- 53: gb\_htg17:\*
- 54: em\_bai:\*
- 55: em\_ba2:\*
- 56: em\_htg1:\*
- 57: em\_htg2:\*
- 58: em\_htg3:\*
- 59: em\_htg4:\*
- 60: em\_htg5:\*
- 61: em\_htg6:\*
- 62: em\_htg7:\*
- 63: em\_hum3:\*
- 64: em\_hum4:\*
- 65: em\_hum5:\*
- 66: em\_hum6:\*
- 67: gb\_htg18:\*
- 68: gb\_htg19:\*
- 69: gb\_htg20:\*
- 70: gb\_htg21:\*
- 71: gb\_htg22:\*
- 72: gb\_htg23:\*
- 73: gb\_htg24:\*
- 74: gb\_htg25:\*
- 75: gb\_htg26:\*
- 76: gb\_htg27:\*
- 77: gb\_htg28:\*
- 78: gb\_htg29:\*
- 79: gb\_htg30:\*
- 80: gb\_htg31:\*
- 81: gb\_vil:\*
- 82: gb\_vil2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	1300.2	68.1	1339	11	AF057557 Homo sapi
2	745	39.0	221365	47	AC023534 Homo sapi
3	160.6	8.4	58628	71	AC027719 Homo sapi
C 4	106	5.5	221365	47	AC023534
5	56.2	2.9	7218	5	I66494 Sequence 14
C 6	55.8	2.9	43147	1	AL109663 Streptomy
C 7	54.2	2.8	55482	44	AC022663 Homo sapi
8	50.4	2.6	227613	68	AC026852 Homo sapi
C 9	49.6	2.6	224850	77	AC022780 Mus muscu
C 10	47.6	2.5	80507	48	AC025066 Homo sapi
11	47.4	2.5	71578	44	AC023440 Homo sapi
12	47.2	2.5	176705	42	AC018348 Homo sapi
13	47	2.5	2904	9	AB005293 Homo sapi
C 14	46.6	2.4	154566	44	AC008334 Drosophil
C 15	46.4	2.4	159758	68	AC020571 Homo sapi
16	45.4	2.4	168012	53	AC025565 Homo sapi
C 17	45.2	2.4	213063	43	AC009964 Homo sapi
C 18	45.2	2.4	219565	43	AC009974 Homo sapi
19	45	2.4	44169	53	AC025851 Homo sapi
C 20	45	2.4	63827	74	AC037448 Homo sapi
21	45	2.4	76968	44	AC022648 Homo sapi
22	45	2.4	82815	78	AC062001 Homo sapi
23	44.8	2.3	10129	2	AE002081 Deinococc
C 24	44.6	2.3	11440	2	AB002034 Deinococc

25 44.6 2.3 136551 76 AC048354 Homo sapi  
26 44.6 2.3 175837 70 AC022045 Homo sapi  
c 27 44.4 2.3 145540 9 AP001052 Homo sapi  
c 28 44.4 2.3 179094 46 AC023429 Homo sapi  
c 29 44.4 2.3 181490 77 AC009690 Homo sapi  
c 30 44.4 2.3 244466 46 AC009235 Homo sapi  
c 31 44.2 2.3 8113 81 HSBICP4A Li4320 Bovine herp  
c 32 44.2 2.3 70542 46 AC022678 Homo sapi  
c 33 44.2 2.3 118831 79 AC064818 Homo sapi  
c 34 44.2 2.3 135301 81 BHV1GEN AC04818 Homo sapi  
c 35 44.2 2.3 135301 81 BHV1GEN AC004801 Bovine he  
c 36 44 2.3 153367 47 AC021096 AC004801 Bovine he  
c 37 43.8 2.3 194828 79 AC006994 AC021096 Homo sapi  
c 38 43.6 2.3 61910 44 AC023002 AC006994 Homo sapi  
c 39 43.6 2.3 78037 74 AC025047 AC023002 Homo sapi  
c 40 43.6 2.3 79264 44 AC022707 AC025047 Mus muscu  
c 41 43.6 2.3 124328 42 AC016442 AC022707 Homo sapi  
c 42 43.6 2.3 236595 48 AC018648 AC016442 Homo sapi  
c 43 43.4 2.3 7218 5 I66494 AC018648 Homo sapi  
c 44 43.4 2.3 65831 76 AC048358 I66494 Sequence 14  
c 45 43.4 2.3 135552 48 AC019251 AC048358 Homo sapi  
AC019251 Homo sapi

ALIGNMENTS

RESULT 1  
AF057557 1339 bp mRNA PRI 04-JUN-1998  
LOCUS Homo sapiens anti-Fas-induced apoptosis (TOSO) mRNA, complete cds.  
DEFINITION AF057557.1 GI:3169292  
ACCESSION AF057557.1  
VERSION 1  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1339)  
AUTHORS Hitoshi,Y., Lorenz,J., Kitada,S.I., Fisher,J., LaBarge,M.,  
Ring,H.Z., Francke,U., Reed,J.C., Kinoshita,S. and Nolan,G.P.  
TITLE Toso, a cell surface, specific regulator of Fas-induced apoptosis  
in T cells  
JOURNAL Immunity 8 (4), 461-471 (1998)  
MEDLINE 98246048  
REFERENCE 2 (bases 1 to 1339)  
AUTHORS LaBarge,M. and Hitoshi,Y.  
TITLE Direct Submission  
JOURNAL Submitted (03-APR-1998) Molecular Pharmacology, Stanford  
University, 300 Pasteur Dr., Palo Alto, CA 94304, USA  
FEATURES  
Location/Qualifiers  
1..1339  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1q31-q32"  
/cell\_line="Jurkat"  
/cell\_type="T cell"  
/note="TOSO.2 cDNA; cloned by PCR using primers designed  
from the original TOSO.1 cDNA sequence"  
1..1339  
/gene="TOSO"  
20..1192  
/gene="TOSO"  
/note="transmembrane protein; Ig superfamily member"  
/codon\_start=1  
/product="anti-Fas-induced apoptosis"  
/protein\_id="AAC18830.1"  
/db\_xref="GI:3169293"  
/translation="MDFWLMPLFLPVSGALRLIPVKVEGLGGSVTIKPLPEMHV  
RIYLICREMGAGTCTVSTNFIKAEYKGRVTLKQYPRKNLFLVEVTLQTFESDGVY  
ACAGNMDRGTOKVLTNVHSEYEPFSEWQPMPTPKWFLPYLPFOMPAVSSKPV  
TRYVTPAQRGVPPVHHSSPTQILTRPVRSSVAGDKPTELPSPATSKLSALEG  
LLAPQTPSYNHHTRLRQRALDVGSGREGQGPHILPILGLFLALLGLVKKRV

ERRKALSRRRRLAVRMRALESSORPRGRPRSONNIYSACPRRARGADAAGTGEAP  
VPCPGAPLPAPLVQSESPWLHAPLSKTSCEVSLYHQPAAMWEDSDDDYINVPA "  
127  
variation /gene="TOSO"  
/note="occurs in TOSO.1 cDNA"  
/replace="a"  
1015  
variation /gene="TOSO"  
/note="occurs in TOSO.1 cDNA"  
/replace="g"  
BASE COUNT 309 a 432 c 340 g 258 t  
ORIGIN  
Query Match 68.1%; Score 1300.2; DB 11; Length 1339;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 1335; Conservative 0; Mismatches 3; Indels 3; Gaps 3;  
QY 55 ttgcactctagaaggacaatggacttctggctttggccactttacttctgcagatc 114  
Db 1 TTGCACCTCTAGAGGGACAATGGACTTCTGGCTTTGGCCACTTTACTTCTGCAGTATC 60  
QY 115 agggccctgaggatcctcccaagtaaggtagagggggagctggcgagatcagttac 174  
Db 61 GGGGCCCTGAGGATCCTCCAGAAAGTAAGGTAGAGGGGAGCTGGGGGATCAGTTAC 120  
QY 175 catcaatgcccacttctgaaatgcatgagagatatctgtgcccggagatggtctg 234  
Db 121 CATCAAGTGCCCACTTCTCGAAATGCATGTGAGGATATATCTGTGCCGGAGATGGCTGG 180  
QY 235 atctggaacatgtgtaccgtggtatccacacaaacttcacaggcagaaacaaagg 294  
Db 181 ATCTGGAACATGTGGTACCGTGGTATCCACCACCACTTCAATCAAGGACGAATACAGGG 240  
QY 295 ccgagttactgaagcaatacccaagcaagaatctgttctcctagtgaggttaacacagt 354  
Db 241 CGGAGTTACTGTGAAGCAATACCCACGCAAGAATCTGTTCCTAGTGGAGGTAAACAGCT 300  
QY 355 gacagaaagtacagcgaggtctatgctcgagcgagcgagcatgaacacagaccgggaaa 414  
Db 301 GACAGAAAGTGACAGCGGAGTCTATGCTCGGAGCGGCATGAAACACACACCGGGGAAA 360  
QY 415 gaccagaaagtcacccctgaaatgtccacagtgaatacagagccatcattgggaagagcgc 474  
Db 361 GACCAGAAAGTCACCCCTGAATGTCCAGTGAATACGAGGCATCATGGGAGAGCAGCC 420  
QY 475 aatgctggaactccaaaatgttccatctgcctattgttccagatgctgcataatgc 534  
Db 421 AATGCTGAGACTCCAAAATGGTTTCATCTGCCCTATTGTTTCCAGATGCTGCATATGC 480  
QY 535 cagttcttccaaattcgttaaccagagttaccacaccagctcaaaaggggcaaggtccctcc 594  
Db 481 CAGTTCTTCCAAAATTCGTAACAGAGTTAGCACACAGCTCAAAAGGGGCAAGTCCCTCC 540  
QY 595 agttcacactctcccccacaccccaaatccaccacccctcagatccagagcagtc 654  
Db 541 AGTTTCAACACTCTCTCCCCACCCCAAAATCAACCCACCCCTCGAGTGTCAGAGCATC 600  
QY 655 ttccagtagcaggtgacaagccccgaaccttctgcatccactacacagctcctcaaaatctc 714  
Db 601 TTTCAGTAGAGGTGACAAAGCCGGAAGCTTCTGTCATCCACTACAGCTCAAAAATCTC 660  
QY 715 agctctgaggggtctcaagccagcagccagctacacacacacacacagaggtgca 774  
Db 661 AGCTCTGGAGGGGCTCTCAGCCCCAGAGCCCCAGCTAGCTACACACACACAGGTCGA 720  
QY 775 caggcagagagcactgactatggtcacagctctggaggggaaggaaggaagatttcaat 834  
Db 721 CAGGCAGAGAGCAGTGGACTATGGCTCAGCTCTGGAGGGAAGGCCAAGGATTTACAT 780  
QY 835 cctaatccccaccactctggccttttctctgctgcaactctggcctggtggtgaaag 894  
Db 781 CCTGATCCGACCATCTCTGGGCGCTTTCTCTGCTGGCATTCTTGGGGCTGTGTGTGAAAG 840



\* 34913 36817: contig of 1905 bp in length  
\* 36818 36917: gap of 100 bp  
\* 36918 38575: contig of 1658 bp in length  
\* 38576 38675: gap of 100 bp  
\* 38676 41309: contig of 2634 bp in length  
\* 41310 41409: gap of 100 bp  
\* 41410 43344: contig of 1935 bp in length  
\* 43345 43444: gap of 100 bp  
\* 43445 45853: contig of 2409 bp in length  
\* 45854 45953: gap of 100 bp  
\* 45954 48434: contig of 2481 bp in length  
\* 48435 48534: gap of 100 bp  
\* 48535 51026: contig of 2492 bp in length  
\* 51027 51126: gap of 100 bp  
\* 51127 54369: contig of 3243 bp in length  
\* 54370 54469: gap of 100 bp  
\* 54470 56661: contig of 2192 bp in length  
\* 56662 56761: gap of 100 bp  
\* 56762 59097: contig of 2336 bp in length  
\* 59098 59197: gap of 100 bp  
\* 59198 61551: contig of 2354 bp in length  
\* 61552 61651: gap of 100 bp  
\* 61652 64257: contig of 2606 bp in length  
\* 64258 64357: gap of 100 bp  
\* 64358 67511: contig of 3154 bp in length  
\* 67512 67611: gap of 100 bp  
\* 67612 70329: contig of 2718 bp in length  
\* 70330 70429: gap of 100 bp  
\* 70430 73572: contig of 3143 bp in length  
\* 73573 73672: gap of 100 bp  
\* 73673 77203: contig of 3531 bp in length  
\* 77204 77303: gap of 100 bp  
\* 77304 80599: contig of 3296 bp in length  
\* 80600 80699: gap of 100 bp  
\* 80700 84046: contig of 3347 bp in length  
\* 84047 84146: gap of 100 bp  
\* 84147 88357: contig of 4211 bp in length  
\* 88358 88457: gap of 100 bp  
\* 88458 92454: contig of 3997 bp in length  
\* 92455 92554: gap of 100 bp  
\* 92555 96892: contig of 4338 bp in length  
\* 96893 96992: gap of 100 bp  
\* 96993 101334: contig of 4342 bp in length  
\* 101335 101434: gap of 100 bp  
\* 101435 105438: contig of 4004 bp in length  
\* 105439 105538: gap of 100 bp  
\* 105539 110413: contig of 4875 bp in length  
\* 110414 110513: gap of 100 bp  
\* 110514 115994: contig of 5481 bp in length  
\* 115995 116094: gap of 100 bp  
\* 116095 120447: contig of 4353 bp in length  
\* 120448 120547: gap of 100 bp  
\* 120548 124483: contig of 3936 bp in length  
\* 124484 124583: gap of 100 bp  
\* 124584 130249: contig of 5666 bp in length  
\* 130250 130349: gap of 100 bp  
\* 130350 135462: contig of 5113 bp in length  
\* 135463 135562: gap of 100 bp  
\* 135563 139847: contig of 4285 bp in length  
\* 139848 139947: gap of 100 bp  
\* 139948 146731: contig of 6784 bp in length  
\* 146732 146831: gap of 100 bp  
\* 146832 151469: contig of 4638 bp in length  
\* 151470 151569: gap of 100 bp  
\* 151570 159901: contig of 8332 bp in length  
\* 159902 160001: gap of 100 bp  
\* 160002 168308: contig of 8307 bp in length  
\* 168309 168408: gap of 100 bp  
\* 168409 176195: contig of 7787 bp in length  
\* 176196 176295: gap of 100 bp  
\* 176296 189303: contig of 13008 bp in length  
\* 189304 189403: gap of 100 bp  
\* 189404 202135: contig of 12732 bp in length

FEATURES  
source

\* 202136 202235: gap of 100 bp  
\* 202236 221365: contig of 19130 bp in length.

Location/Qualifiers  
1. .221365  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="RPCI-II Human Male BAC"  
1. .11108  
/note="assembly\_fragment"  
1209. .2292  
/note="assembly\_fragment"  
2393. .3673  
/note="assembly\_fragment"  
3774. .4947  
/note="assembly\_fragment"  
5048. .6104  
/note="assembly\_fragment"  
6205. .7259  
/note="assembly\_fragment"  
7360. .8667

Query Match 39.0%; Score 745; DB 47; Length 221365;

Best Local Similarity 98.3%; Pred. No. 7.1e-178;

Matches 785; Conservative 0; Mismatches 10; Indels 4; Gaps 3;

QY 1114 caagatctgaatctccctgagctccatcccccatctctaaagaccagctgtaacatg 1173  
|||||  
DB 58030 CAGGTGTCGTAATCTCCCTGGCTCCATGCCCTCTCTGAAGACCAGCTGGAATACGG 58089  
QY 1174 agccttaccaccagcctgcgcgcctgagtgagacagtgattcagatgactacatcaat 1233  
|||||  
DB 58090 AGCCTTACCACCAGCCTGCCCATGATGAGGACACAGTGAATTCAGATGACTACATCAAT 58149  
QY 1234 gtctcgcctgacaaactcccagctatcccccaaccccagctcgagctgggtgccaag 1293  
|||||  
DB 58150 GTTCTGCTGACACTCCCCAGCTATCCCCCAACCCAGGCTCGGACTGTGGTCCAAAG 58209  
QY 1294 gagtcctcatctctgctgagtgcccaatacctgctcatgtgttctcagagccctcatca 1353  
|||||  
DB 58210 GAGTCTATCTATCTGCTGATGTCCTCACTGCTTCTGCTGCTGCTGCTGCTGCTGCTG 58269  
QY 1354 ctctccatgcccatctctgactcccatcccatctctctgtgcccctgagcagctgctgt 1413  
|||||  
DB 58270 CTTCCTCATGCCCATCTCGACTCCCATCCCATCTATCTGT-GCCCTGAGCATGGCTCTG 58328  
QY 1414 cccccaggtcgtcttgacaccccttgacagcccccctgtagttgacaggtgaagctgtagcca 1473  
|||||  
DB 58329 CCCCCAGGTGCTTTGACACACCTTGGCAGCCCCCTGTAGTTGACAGGTAACTGTAGGCA 58388  
QY 1474 tgttagcgaattgtcccaatgccattgctcttcccaagcgtcgacagactgtggg 1533  
|||||  
DB 58389 TGTAGAGCAATGTGCCAATGCCATGTGCTTCTTCCAAAGCCGTCGAACAGCTGTGGG 58448  
QY 1534 attgacagagtgttcttccatgtctttgacacaggggtgtgtgtgctgcagagctctag 1593  
|||||  
DB 58449 ATTTGACAGAGTGTCTTCTTCCATGTCTTTGACACAGGGGT-TGTTGCTGCCCAGGCTCTAG 58507  
QY 1594 atcaatggcagcagctggggcagagggcagatgattctctcgggcatcttcccaggg 1653  
|||||  
DB 58508 ATCATATGTCATCAGGCTGGGGCAGAGGATAGCTATTGCTCGGGGATCTCTCCAGGG 58567  
QY 1654 ttgggtcttacacaaatagaaggctcttctgagttatgagcgtgctcagagcccat 1713  
|||||  
DB 58568 TTGGGTCTTACACAAATAGAGGCTCTTGTCTCTGAGTTATGTGATGATGCCCTAGCCCCAT 58627  
QY 1714 ggactaagcaggggtctgggtat--aaacactcctggaaacgcttttgcctgtatccaaat 1771  
|||||  
DB 58628 GGACTAAGCAGGGGGTCTGGTATATAAAACACTCCTGTGAAACGCCCTTTGCTGCTGATCAAAAT 58687  
QY 1772 gtagcactgtctaagcagctcactctatctcagttctatgctaaagcgaatttatct 1831  
|||||  
DB 58688 GTTAGCACTTGCTAGTGAAGCTCTACTTATCTCAAGTTCTTAGCTTAAGGCAATTATCT 58747

```

Oy 1832 tgatgtgataaaccacactattagcaagatatgcataatctctt 1891
      |||||||
Db 58748 TGATGTGATGATAAACCAACTATTATTAGCAAGATATGCATATATATCCATAAATCTCTT 58807
      |||||||
Oy 1892 tactctgtctccatctt 1910
      |||||||
Db 58808 TACTCTGTCGCCACTT 58826
      |||||||

RESULT 3
AC027719/c
LOCUS      58628 bp      DNA      HTG      01-APR-2000
DEFINITION Homo sapiens chromosome 8 clone RP11-258P9 map 8, LOW-PASS SEQUENCE
            SAMPLING.
ACCESSION AC027719
VERSION    AC027719.1 GI:7382531
KEYWORDS   HTG; HTGS_PHASE0.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 58628)
            Birren, B., Linton, L., Nusbaum, C. and Lander, E.
            Homo sapiens chromosome 8, clone RP11-258P9
            Unpublished
            2 (bases 1 to 58628)
            Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
            Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bada, F.,
            Boguslavsky, L., Bouckhalter, B., Brown, A., Burkett, G.,
            Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
            Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,
            Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
            Galegher, J., Gardyna, S., Glaser, S., Goyette, M., Graham, L.,
            Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
            Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
            Klein, J., Lacroque, K., Lamazares, R., Landers, T., Lehoczy, J.,
            Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
            McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
            Melchior, J., Menes, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
            Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
            O'Neill, B., Oliver, J.M., Oliver, J., Peterson, K., Pierre, N.,
            Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
            Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
            Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
            Testfaye, S., Theodore, J., Tirtell, A., Travers, M., Trigilio, J.,
            Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
            Young, G., Zainoun, J., Zimmer, A. and Zody, M.
            Direct Submission
            Submitted (01-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ----- Project Information
            Center project name: L7449
            Center clone name: 258_P_9
            -----
            * NOTE: This record contains 74 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will

```

```

* 28474 29165: contig of 692 bp in length
* 29166 29265: gap of 100 bp
* 29266 29365: contig of 691 bp in length
* 29365 30056: gap of 100 bp
* 30057 30741: contig of 685 bp in length
* 30742 30841: gap of 100 bp
* 30842 31519: contig of 678 bp in length
* 31520 31619: gap of 100 bp
* 31620 32323: contig of 704 bp in length
* 32324 32423: gap of 100 bp
* 32424 33110: contig of 687 bp in length
* 33111 33210: gap of 100 bp
* 33211 33902: contig of 692 bp in length
* 33903 34002: gap of 100 bp
* 34003 34708: contig of 706 bp in length
* 34709 34808: gap of 100 bp
* 34809 35500: contig of 692 bp in length
* 35501 35600: gap of 100 bp
* 35601 36300: contig of 700 bp in length
* 36301 36400: gap of 100 bp
* 36401 37097: contig of 697 bp in length
* 37098 37197: gap of 100 bp
* 37198 37900: contig of 703 bp in length
* 37901 38000: gap of 100 bp
* 38001 38695: contig of 695 bp in length
* 38696 38795: gap of 100 bp
* 38796 39487: contig of 692 bp in length
* 39488 39587: gap of 100 bp
* 39588 40291: contig of 704 bp in length
* 40292 40391: gap of 100 bp
* 40392 41101: contig of 710 bp in length
* 41102 41201: gap of 100 bp
* 41202 41886: contig of 685 bp in length
* 41887 41986: gap of 100 bp
* 41987 42686: contig of 700 bp in length
* 42687 42786: gap of 100 bp
* 42787 43484: contig of 698 bp in length
* 43485 43584: gap of 100 bp
* 43585 44276: contig of 692 bp in length
* 44277 44376: gap of 100 bp
* 44377 45078: contig of 702 bp in length
* 45079 45178: gap of 100 bp
* 45179 45869: contig of 691 bp in length
* 45870 45969: gap of 100 bp
* 45970 46664: contig of 695 bp in length
* 46665 46764: gap of 100 bp
* 46765 47456: contig of 692 bp in length
* 47457 47556: gap of 100 bp
* 47557 48249: contig of 693 bp in length
* 48250 48349: gap of 100 bp
* 48350 49042: contig of 693 bp in length
* 49043 49142: gap of 100 bp
* 49143 49843: contig of 701 bp in length
* 49844 49943: gap of 100 bp
* 49944 50640: contig of 697 bp in length
* 50641 50740: gap of 100 bp
* 50741 51438: contig of 698 bp in length
* 51439 51538: gap of 100 bp
* 51539 52231: contig of 693 bp in length
* 52232 52331: gap of 100 bp
* 52332 53020: contig of 689 bp in length
* 53021 53120: gap of 100 bp
* 53121 53857: contig of 737 bp in length
* 53858 53957: gap of 100 bp
* 53958 54652: contig of 695 bp in length
* 54653 54752: gap of 100 bp

Query Match      8.4%; Score 160.6; DB 71: Length 58628;
Best Local Similarity 97.6%; Pred. No. 6e-30;
Matches 163; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 102 tccgtccatccagggccctgaggtccctccagagtaagtagagggagctgg 161
|||||

```

```

Db 40603 TCCCTGCAGTATCAGGGCCCTCAGGATCCTCCAGAGTAAGTAGAGGGGAGCTGG 40544
Qy 162 gggatcagttaccatcaaatgccacattcctgaaatgcatgtgagatatctgtgcc 221
|||||
Db 40543 CGCGATCAGTTACCATCAAGTCCGACCTTCCTGAAATGCGATGTGAGGATATATCTGTGCC 40484
|||||
Qy 222 gggagatggtgagatctggaacatggtgacgtggtatccaccacc 268
|||||
Db 40483 GGGAGATGCGTGGATCTGGAACATGTGGTACCGTGTATCCACCACC 40437
|||||

RESULT 4
AC023534/2
LOCUS AC023534 221365 bp DNA HTG 02-MAR-2000
DEFINITION Homo sapiens clone RP11-462N18, *** SEQUENCING IN PROGRESS ***, 61
unordered pieces.
ACCESSION AC023534
VERSION AC023534.2 GI:7143452
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 221365)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome, clone RP11-462N18
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 221365)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,
Bouckgaert,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepe,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,
DeArelano,K., Dewar,K., Dodge,S., Domingo,M., Doyle,M.,
Festor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
Galagan,P., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Landers,T., Larcocque,K., Lehoczy,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrum,J.,
Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,T.M.,
Peterson,K., Pierre,N., Pisan,L., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tirrell,A.,
Travers,M., Triglio,J., Vassiliev,H., Viel,R., Vo,A., Willson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
Zody,M.
Direct Submission
TITLE Submitted (15-FEB-2000) Whitehead Institute/MIT Center for Genome
JOURNAL Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Mar 2, 2000 this sequence version replaced gi:6978210.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5244
Center clone name: 462_N_18
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 61 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```



\* 1 1108: contig of 1108 bp in length  
\* 1109 1208: gap of 100 bp  
\* 1209 2292: contig of 1084 bp in length  
\* 2293 2392: gap of 100 bp  
\* 2393 3673: contig of 1281 bp in length  
\* 3674 3773: gap of 100 bp  
\* 3774 4947: contig of 1174 bp in length  
\* 4948 5047: gap of 100 bp  
\* 5048 6104: contig of 1057 bp in length  
\* 6105 6204: gap of 100 bp  
\* 6205 7259: contig of 1055 bp in length  
\* 7260 7359: gap of 100 bp  
\* 7360 8667: contig of 1308 bp in length  
\* 8668 8767: gap of 100 bp  
\* 8768 10566: contig of 1799 bp in length  
\* 10567 10666: gap of 100 bp  
\* 10667 12173: contig of 1507 bp in length  
\* 12174 12273: gap of 100 bp  
\* 12274 13643: contig of 1370 bp in length  
\* 13644 13743: gap of 100 bp  
\* 13744 15215: contig of 1472 bp in length  
\* 15216 15315: gap of 100 bp  
\* 15316 16609: contig of 1294 bp in length  
\* 16610 16709: gap of 100 bp  
\* 16710 18085: contig of 1376 bp in length  
\* 18086 18185: gap of 100 bp  
\* 18186 19668: contig of 1483 bp in length  
\* 19669 19768: gap of 100 bp  
\* 19769 22083: contig of 2315 bp in length  
\* 22084 22183: gap of 100 bp  
\* 22184 24052: contig of 1869 bp in length  
\* 24053 24152: gap of 100 bp  
\* 24153 25902: contig of 1750 bp in length  
\* 25903 26002: gap of 100 bp  
\* 26003 27639: contig of 1637 bp in length  
\* 27640 27739: gap of 100 bp  
\* 27740 29377: contig of 1638 bp in length  
\* 29378 29477: gap of 100 bp  
\* 29478 30910: contig of 1433 bp in length  
\* 30911 31010: gap of 100 bp  
\* 31011 31225: contig of 215 bp in length  
\* 31226 31325: gap of 100 bp  
\* 31326 32890: contig of 1565 bp in length  
\* 32891 32990: gap of 100 bp  
\* 32991 34812: contig of 1822 bp in length  
\* 34813 34912: gap of 100 bp  
\* 34913 36817: contig of 1905 bp in length  
\* 36818 36917: gap of 100 bp  
\* 36918 38575: contig of 1658 bp in length  
\* 38576 38675: gap of 100 bp  
\* 38676 41309: contig of 2634 bp in length  
\* 41310 41409: gap of 100 bp  
\* 41410 43344: contig of 1935 bp in length  
\* 43345 43444: gap of 100 bp  
\* 43445 45853: contig of 2409 bp in length  
\* 45854 45953: gap of 100 bp  
\* 45954 48434: contig of 2481 bp in length  
\* 48435 48534: gap of 100 bp  
\* 48535 51026: contig of 2492 bp in length  
\* 51027 51126: gap of 100 bp  
\* 51127 54369: contig of 3243 bp in length  
\* 54370 54469: gap of 100 bp  
\* 54470 56661: contig of 2192 bp in length  
\* 56662 56761: gap of 100 bp  
\* 56762 59097: contig of 2336 bp in length  
\* 59098 59197: gap of 100 bp  
\* 59198 61551: contig of 2354 bp in length  
\* 61552 61651: gap of 100 bp  
\* 61652 64257: contig of 2606 bp in length  
\* 64258 64357: gap of 100 bp  
\* 64358 67511: contig of 3154 bp in length  
\* 67512 67611: gap of 100 bp  
\* 67612 70329: contig of 2718 bp in length

\* 70330 70429: gap of 100 bp  
\* 70430 73572: contig of 3143 bp in length  
\* 73573 73672: gap of 100 bp  
\* 73673 77203: contig of 3531 bp in length  
\* 77204 77303: gap of 100 bp  
\* 77304 80599: contig of 3296 bp in length  
\* 80600 80699: gap of 100 bp  
\* 80700 84046: contig of 3347 bp in length  
\* 84047 84146: gap of 100 bp  
\* 84147 88357: contig of 4211 bp in length  
\* 88358 88457: gap of 100 bp  
\* 88458 92454: contig of 3997 bp in length  
\* 92455 92554: gap of 100 bp  
\* 92555 96892: contig of 4338 bp in length  
\* 96893 96992: gap of 100 bp  
\* 96993 101334: contig of 4342 bp in length  
\* 101335 101434: gap of 100 bp  
\* 101435 105438: contig of 4004 bp in length  
\* 105439 105538: gap of 100 bp  
\* 105539 110413: contig of 4875 bp in length  
\* 110414 110513: gap of 100 bp  
\* 110514 115994: contig of 5481 bp in length  
\* 115995 116094: gap of 100 bp  
\* 116095 120447: contig of 4353 bp in length  
\* 120448 120547: gap of 100 bp  
\* 120548 124483: contig of 3936 bp in length  
\* 124484 124583: gap of 100 bp  
\* 124584 130249: contig of 5666 bp in length  
\* 130250 130349: gap of 100 bp  
\* 130350 135462: contig of 5113 bp in length  
\* 135463 135562: gap of 100 bp  
\* 135563 139847: contig of 4285 bp in length  
\* 139848 139947: gap of 100 bp  
\* 139948 146731: contig of 6784 bp in length  
\* 146732 146831: gap of 100 bp  
\* 146832 151469: contig of 4638 bp in length  
\* 151470 151569: gap of 100 bp  
\* 151570 159901: contig of 8332 bp in length  
\* 159902 160001: gap of 100 bp  
\* 160002 168308: contig of 8307 bp in length  
\* 168309 168408: gap of 100 bp  
\* 168409 176195: contig of 7787 bp in length  
\* 176196 176295: gap of 100 bp  
\* 176296 189303: contig of 13008 bp in length  
\* 189304 189403: gap of 100 bp  
\* 189404 202135: contig of 12732 bp in length  
\* 202136 202235: gap of 100 bp  
\* 202236 221365: contig of 19130 bp in length.

FEATURES

source

1. 221365  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="RP11-462M18"

/clone\_lib="RPC1-11 Human Male BAC"  
1. 11108  
/note="assembly\_fragment"

misc\_feature

1209..2292  
/note="assembly\_fragment"

misc\_feature

2393..3673  
/note="assembly\_fragment"

misc\_feature

3774..4947  
/note="assembly\_fragment"

misc\_feature

5048..6104  
/note="assembly\_fragment"

misc\_feature

6205..7259  
/note="assembly\_fragment"

misc\_feature

7360..8667  
/note="assembly\_fragment"

Query Match

Best Local Similarity 5.5%; Score 106; DB 47; Length 221365;

Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 gagtaagcagcgtgtctccatcccccctctctagggggtctgttgatggaccttgcaactcta 64





```

LOCUS       AC022663       55482 bp      DNA           HTG           06-FEB-2000
DEFINITION  Homo sapiens clone RP11-2905, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC022663
VERSION     AC022663.1 GI:6910643
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 55482)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens, clone RP11-2905
JOURNAL
REFERENCE   2 (bases 1 to 55482)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Bada,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepe,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeArelano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Hearford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,Y., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Schojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smitt, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L4854
Center clone name: 29_Q_5
-----
* NOTE: This record contains 77 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1 734: contig of 734 bp in length
1 735 gap of unknown length
1 736 contig of 722 bp in length
1 1457 gap of unknown length
1 2153 contig of 696 bp in length
1 2848 contig of 696 bp in length
1 2849 gap of unknown length
1 3545 contig of 697 bp in length
1 3546 gap of unknown length
1 4229 contig of 683 bp in length
1 4229 gap of unknown length
1 4948 contig of 720 bp in length
1 4949 gap of unknown length
1 5662 contig of 714 bp in length
1 5662 gap of unknown length
1 5663 contig of 709 bp in length
1 5663 gap of unknown length
1 6372 contig of 726 bp in length
1 6372 gap of unknown length
1 7098 contig of 738 bp in length
1 7098 gap of unknown length
1 7836 contig of 729 bp in length
1 7836 gap of unknown length
1 8565 contig of 751 bp in length
1 8565 gap of unknown length
1 9316 contig of 814 bp in length
1 9316 gap of unknown length
1 10130 contig of 765 bp in length
1 10130 gap of unknown length
1 10895 contig of 782 bp in length
1 10895 gap of unknown length
1 11677 contig of 715 bp in length
1 11677 gap of unknown length
1 12392 contig of 706 bp in length
1 12392 gap of unknown length
1 13098 contig of 710 bp in length
1 13098 gap of unknown length
1 13808 contig of 725 bp in length
1 13808 gap of unknown length
1 14533 contig of 736 bp in length
1 14533 gap of unknown length
1 15269 contig of 737 bp in length
1 15269 gap of unknown length
1 16006 contig of 683 bp in length
1 16006 gap of unknown length
1 16689 contig of 714 bp in length
1 16689 gap of unknown length
1 17403 contig of 714 bp in length
1 17403 gap of unknown length
1 18117 contig of 714 bp in length
1 18117 gap of unknown length
1 18831 contig of 713 bp in length
1 18831 gap of unknown length
1 19544 contig of 707 bp in length
1 19544 gap of unknown length
1 20251 contig of 735 bp in length
1 20251 gap of unknown length
1 20986 contig of 721 bp in length
1 20986 gap of unknown length
1 21707 contig of 741 bp in length
1 21707 gap of unknown length
1 22448 contig of 729 bp in length
1 22448 gap of unknown length
1 23177 contig of 731 bp in length
1 23177 gap of unknown length
1 23908 contig of 719 bp in length
1 23908 gap of unknown length
1 24627 contig of 711 bp in length
1 24627 gap of unknown length
1 25338 contig of 722 bp in length
1 25338 gap of unknown length
1 26060 contig of 715 bp in length
1 26060 gap of unknown length
1 26775 contig of 710 bp in length
1 26775 gap of unknown length
1 27485 contig of 727 bp in length
1 27485 gap of unknown length
1 28212 contig of 725 bp in length
1 28212 gap of unknown length
1 28937 contig of 685 bp in length
1 28937 gap of unknown length
1 29622 contig of 723 bp in length
1 29622 gap of unknown length
1 30345 contig of 709 bp in length
1 30345 gap of unknown length
1 31054 contig of 716 bp in length
1 31054 gap of unknown length
1 31770 contig of 709 bp in length
1 31770 gap of unknown length

```







	* 196773	196872:	gap of	100 bp	
	*	196873	224850:	contig of 27978 bp in length.	
FEATURES	Location/Qualifiers				
source	1..224850				
	/organism=	"Mus musculus"			
	/db_xref=taxon:	10090"			
	/chromosome=	"11"			
	/map=	"11"			
	/clone=	"RP23-179C3"			
	/clone_lib=	"RPCI-23 Female Mouse BAC"			
misc_feature	1..1484				
	/note=	"assembly_fragment"			
misc_feature	1585..3154				
	/note=	"assembly_fragment"			
misc_feature	3255..4975				
	/note=	"assembly_fragment"			
misc_feature	5076..7040				
	/note=	"assembly_fragment"			
misc_feature	7141..8876				
	/note=	"assembly_fragment"			
misc_feature	8977..10661				
	/note=	"assembly_fragment"			
misc_feature	10762..13062				
	/note=	"assembly_fragment"			
misc_feature	13163..15458				
	/note=	"assembly_fragment"			
misc_feature	15559..18599				
	/note=	"assembly_fragment"			
misc_feature	18700..21120				
	/note=	"assembly_fragment"			
misc_feature	21221..23597				
	/note=	"assembly_fragment"			
misc_feature	23698..26190				
	/note=	"assembly_fragment"			
misc_feature	26291..29227				
	/note=	"assembly_fragment"			
misc_feature	29328..31847				
	/note=	"assembly_fragment"			
misc_feature	31948..35346				
	/note=	"assembly_fragment"			
	/clone_end=T7				
	/vector_side:right				
misc_feature	35447..40435				
	/note=	"assembly_fragment"			
misc_feature	40536..44230				
	/note=	"assembly_fragment"			
misc_feature	44331..49897				
	/note=	"assembly_fragment"			
misc_feature	49998..54610				
	/note=	"assembly_fragment"			
misc_feature	54711..59416				
	/note=	"assembly_fragment"			
misc_feature	59517..64639				
	/note=	"assembly_fragment"			
misc_feature	64740..69945				
	/note=	"assembly_fragment"			
misc_feature	70046..75794				
	/note=	"assembly_fragment"			
	/clone_end:Sf6				
	/vector_side:left				
misc_feature	75895..82836				
	/note=	"assembly_fragment"			
misc_feature	82937..89775				
	/note=	"assembly_fragment"			
<hr/>					
Query Match	2.6%; Score 49.6; DB 77; Length 224850;				
Best local similarity	52.4%; Pred. No. 0.083;				
Matches 109;	Conservative	0;	Mismatches	99;	Gaps 0;
<hr/>					



\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 762: contig of 762 bp in length  
763 862: gap of 100 bp  
863 1656: contig of 794 bp in length  
1657 1756: gap of 100 bp  
1757 2583: contig of 807 bp in length  
2584 2663: gap of 100 bp  
2664 3449: contig of 786 bp in length  
3450 3549: gap of 100 bp  
3550 4325: contig of 776 bp in length  
4326 4425: gap of 100 bp  
4426 5227: contig of 802 bp in length  
5228 5327: gap of 100 bp  
5328 6139: contig of 812 bp in length  
6140 6239: gap of 100 bp  
6240 7044: contig of 805 bp in length  
7045 7144: gap of 100 bp  
7145 7849: contig of 805 bp in length  
7850 8049: gap of 100 bp  
8050 8868: contig of 819 bp in length  
8869 8968: gap of 100 bp  
8969 9801: contig of 833 bp in length  
9802 9901: gap of 100 bp  
9902 10610: contig of 703 bp in length  
10611 10710: gap of 100 bp  
10711 11507: contig of 797 bp in length  
11508 11607: gap of 100 bp  
11608 12409: contig of 802 bp in length  
12410 12509: gap of 100 bp  
12510 13307: contig of 798 bp in length  
13308 13407: gap of 100 bp  
13408 14187: contig of 780 bp in length  
14188 14287: gap of 100 bp  
14288 15062: contig of 775 bp in length  
15063 15162: gap of 100 bp  
15163 15978: contig of 816 bp in length  
15979 16078: gap of 100 bp  
16079 16875: contig of 797 bp in length  
16876 16975: gap of 100 bp  
16976 17676: contig of 701 bp in length  
17677 17776: gap of 100 bp  
17777 18581: contig of 805 bp in length  
18582 18681: gap of 100 bp  
18682 19442: contig of 761 bp in length  
19443 19542: gap of 100 bp  
19543 20326: contig of 784 bp in length  
20327 20426: gap of 100 bp  
20427 21241: contig of 815 bp in length  
21242 21341: gap of 100 bp  
21342 22132: contig of 791 bp in length  
22133 22232: gap of 100 bp  
22233 23027: contig of 795 bp in length  
23028 23127: gap of 100 bp  
23128 23916: contig of 789 bp in length  
23917 24016: gap of 100 bp  
24017 24826: contig of 810 bp in length  
24827 24926: gap of 100 bp  
24927 25738: contig of 812 bp in length  
25739 25838: gap of 100 bp  
25839 26646: contig of 808 bp in length  
26647 26746: gap of 100 bp  
26747 27556: contig of 810 bp in length  
27557 27656: gap of 100 bp  
27657 28460: contig of 804 bp in length  
28461 28560: gap of 100 bp  
28561 29364: contig of 804 bp in length  
29365 29464: gap of 100 bp  
29465 30259: contig of 795 bp in length  
30260 30359: gap of 100 bp

30360 31149: contig of 790 bp in length  
31150 31249: gap of 100 bp  
31250 32058: contig of 809 bp in length  
32059 32158: gap of 100 bp  
32159 32968: contig of 810 bp in length  
32969 33068: gap of 100 bp  
33069 33870: contig of 802 bp in length  
33871 33970: gap of 100 bp  
33971 34769: contig of 799 bp in length  
34770 34869: gap of 100 bp  
34870 35682: contig of 813 bp in length  
35683 35782: gap of 100 bp  
35783 36585: contig of 803 bp in length  
36586 36685: gap of 100 bp  
36686 37501: contig of 816 bp in length  
37502 37601: gap of 100 bp  
37602 38393: contig of 792 bp in length  
38394 38493: gap of 100 bp  
38494 39303: contig of 810 bp in length  
39304 39403: gap of 100 bp  
39404 40198: contig of 795 bp in length  
40199 40298: gap of 100 bp  
40299 41092: contig of 794 bp in length  
41093 41192: gap of 100 bp  
41193 41982: contig of 790 bp in length  
41983 42082: gap of 100 bp  
42083 42908: contig of 826 bp in length  
42909 43008: gap of 100 bp  
43009 43788: contig of 780 bp in length  
43789 43888: gap of 100 bp  
43889 44662: contig of 774 bp in length  
44663 44762: gap of 100 bp  
44763 45556: contig of 794 bp in length  
45557 45656: gap of 100 bp  
45657 46462: contig of 806 bp in length  
46463 46562: gap of 100 bp  
46563 47367: contig of 805 bp in length  
47368 47467: gap of 100 bp  
47468 48267: contig of 800 bp in length  
48268 48367: gap of 100 bp  
48368 49156: contig of 789 bp in length  
49157 49256: gap of 100 bp  
49257 50060: contig of 804 bp in length  
50061 50160: gap of 100 bp  
50161 50952: contig of 792 bp in length  
50953 51052: gap of 100 bp  
51053 51860: contig of 808 bp in length  
51861 51960: gap of 100 bp  
51961 52753: contig of 793 bp in length  
52754 52853: gap of 100 bp  
52854 53650: contig of 797 bp in length  
53651 53750: gap of 100 bp  
53751 54532: contig of 782 bp in length  
54533 54632: gap of 100 bp  
54633 55430: contig of 798 bp in length  
55431 55530: gap of 100 bp  
55531 56307: contig of 777 bp in length  
56308 56407: gap of 100 bp  
56408 57177: contig of 770 bp in length  
57178 57277: gap of 100 bp  
57278 57994: contig of 717 bp in length  
57995 58094: gap of 100 bp  
58095 58864: contig of 770 bp in length  
58865 58964: gap of 100 bp  
58965 59764: contig of 800 bp in length  
59765 59864: gap of 100 bp  
59865 60654: contig of 790 bp in length  
60655 60754: gap of 100 bp  
60755 61549: contig of 795 bp in length

Query Match 2.5%; Score 47.6; DB 48; Length 80507;  
Best Local Similarity 49.2%; Pred. No. 0.25;  
Matches 122; Conservative 0; Mismatches 126; Indels 0; Gaps 0;











```
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0387H05
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 78 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1287: contig of 1287 bp in length
*
1288 1406: gap of unknown length
*
1407 2639: contig of 1233 bp in length
*
2640 2758: gap of unknown length
*
2759 3949: contig of 1191 bp in length
*
3950 4068: gap of unknown length
*
4069 6120: contig of 2052 bp in length
*
6121 6239: gap of unknown length
*
6240 7854: contig of 1615 bp in length
*
7855 7973: gap of unknown length
*
7974 9166: contig of 1193 bp in length
*
9167 9285: gap of unknown length
*
9286 10469: contig of 1184 bp in length
*
10470 10588: gap of unknown length
*
10589 12886: contig of 2298 bp in length
*
12887 13005: gap of unknown length
*
13006 14222: contig of 1217 bp in length
*
14223 14341: gap of unknown length
*
14342 16064: contig of 1723 bp in length
*
16065 16183: gap of unknown length
*
16184 18245: contig of 2062 bp in length
*
18246 18364: gap of unknown length
*
18365 19779: contig of 1415 bp in length
*
19780 19898: gap of unknown length
*
19899 21787: contig of 1889 bp in length
*
21788 21906: gap of unknown length
*
21907 23682: contig of 1776 bp in length
*
23683 23801: gap of unknown length
*
23802 25816: contig of 2015 bp in length
*
25817 25935: gap of unknown length
*
25936 27544: contig of 1609 bp in length
*
27545 27663: gap of unknown length
*
27664 29356: contig of 1693 bp in length
*
29357 29475: gap of unknown length
*
29476 31588: contig of 2113 bp in length
*
31589 31707: gap of unknown length
*
31708 33985: contig of 2278 bp in length
*
33986 34104: gap of unknown length
*
34105 35708: contig of 1604 bp in length
*
35709 35827: gap of unknown length
*
35828 38639: contig of 2812 bp in length
*
38640 38758: gap of unknown length
*
38759 40502: contig of 1744 bp in length
*
40503 40621: gap of unknown length
*
40622 42742: contig of 2121 bp in length
*
42743 42861: gap of unknown length
*
42862 44999: contig of 2138 bp in length
*
45000 45118: gap of unknown length
*
45119 46962: contig of 1844 bp in length
*
46963 47081: gap of unknown length
*
47082 49731: contig of 2650 bp in length
*
49732 49850: gap of unknown length
*
49851 51830: contig of 1980 bp in length
*
51831 51949: gap of unknown length
*
51950 55175: contig of 3226 bp in length
*
55176 55294: gap of unknown length
*
55295 59415: contig of 4121 bp in length
*
```

```
*
59416 59534: gap of unknown length
*
59535 62555: contig of 3021 bp in length
*
62556 62674: gap of unknown length
*
62675 68133: contig of 5459 bp in length
*
68134 68252: gap of unknown length
*
68253 72424: contig of 4172 bp in length
*
72425 72543: gap of unknown length
*
72544 77507: contig of 4964 bp in length
*
77508 77625: gap of unknown length
*
77626 83661: contig of 6036 bp in length
*
83662 83779: gap of unknown length
*
83780 90116: contig of 6337 bp in length
*
90117 90234: gap of unknown length
*
90235 96466: contig of 6232 bp in length
*
96467 96584: gap of unknown length
*
96585 102699: contig of 6115 bp in length
*
102700 102817: gap of unknown length
*
102818 103977: contig of 1160 bp in length
*
103978 104095: gap of unknown length
*
104096 105583: contig of 1488 bp in length
*
105584 105701: gap of unknown length
*
105702 106817: contig of 1116 bp in length
*
106818 106935: gap of unknown length
*
106936 108385: contig of 1450 bp in length
*
108386 108503: gap of unknown length
*
108504 109623: contig of 1120 bp in length
*
109624 109741: gap of unknown length
*
109742 110900: contig of 1159 bp in length
*
110901 111018: gap of unknown length
*
111019 112324: contig of 1306 bp in length
*
112325 112442: gap of unknown length
*
112443 113479: contig of 1037 bp in length
*
113480 113597: gap of unknown length
*
113598 114717: contig of 1120 bp in length
*
114718 114835: gap of unknown length
*
114836 115955: contig of 1120 bp in length
*
115956 116073: gap of unknown length
*
116074 117082: contig of 1009 bp in length
*
117083 117200: gap of unknown length
*
117201 118249: contig of 1049 bp in length
*
118250 118367: gap of unknown length
*
118368 119584: contig of 1217 bp in length
*
119585 119702: gap of unknown length
*
119703 120976: contig of 1274 bp in length
*
120977 121094: gap of unknown length
*
121095 122261: contig of 1166 bp in length
*
122262 122378: gap of unknown length
*
122379 123710: contig of 1331 bp in length
*
123710 123827: gap of unknown length
*
123828 125046: contig of 1219 bp in length
*
125047 125164: gap of unknown length
*
125165 126183: contig of 1019 bp in length
*
126184 126301: gap of unknown length
*
126302 127831: contig of 1530 bp in length
*
127832 127949: gap of unknown length
*
127950 128984: contig of 1035 bp in length
*
128985 129102: gap of unknown length
*
129103 130310: contig of 1208 bp in length
*
130311 130428: gap of unknown length
*
130429 131514: contig of 1086 bp in length
*
131515 131632: gap of unknown length
*
131633 132732: contig of 1100 bp in length
*
132733 132850: gap of unknown length
*
132851 134425: contig of 1575 bp in length
*
134426 134543: gap of unknown length
*
134544 135606: contig of 1063 bp in length
*
135607 135724: gap of unknown length
*
135725 137036: contig of 1312 bp in length
*
137037 137154: gap of unknown length
*
137155 138607: contig of 1453 bp in length
*
138608 138725: gap of unknown length
*
138726 139899: contig of 1174 bp in length
*
139900 140017: gap of unknown length
*
```

Tue Sep 12 15:26:52 2000

```
* 140018 141082: contig of 1065 bp in length
* 141083 141200: gap of unknown length
* 141201 142632: contig of 1432 bp in length
* 142633 142750: gap of unknown length
* 142751 143769: contig of 1019 bp in length
* 143770 143887: gap of unknown length
* 143888 144943: contig of 1056 bp in length
* 144944 145061: gap of unknown length
* 145062 146194: contig of 1133 bp in length
* 146195 146312: gap of unknown length
* 146313 148145: contig of 1833 bp in length
* 148146 148263: gap of unknown length
* 148264 150560: contig of 2297 bp in length
* 150561 150678: gap of unknown length
* 150679 152449: contig of 1771 bp in length
* 152450 152567: gap of unknown length
* 152568 153960: contig of 1393 bp in length
* 153961 154078: gap of unknown length
* 154079 155111: contig of 1033 bp in length
* 155112 155229: gap of unknown length
* 155230 156212: contig of 983 bp in length
* 156213 156330: gap of unknown length
* 156331 157331: contig of 1001 bp in length
* 157332 157449: gap of unknown length
* 157450 159758: contig of 2309 bp in length.
FEATURES             Location/Qualifiers
     source           1..159758
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-387H5"
BASE COUNT   42756 a 32163 c 33088 g 41008 t 10743 others
ORIGIN
```

```
Query Match      2.4%; Score 46.4; DB 68; Length 159758;
Best Local Similarity 47.1%; Pred. No. 0.53;
Matches 128; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

QY  926  cgggcccgcgactgacctgagatgcgccttgagagctcccagagcggccggcg 985
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 144889 CGNCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 144830

QY  986  tcgacgcgacgcgctcccaaaacatctacagcgcctgcgcgagcggtcttgag 1045
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 144829 CCGCGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 144770

QY  1046  cggacgctcagggacagggggggcccgctcccgcccgagcgccgttgccccc 1105
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 144769 GGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 144710

QY  1106  ccccgctcaggtgtctgaatctcctgagctccatgccccatctctgaagaccagctgtg 1165
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 144709 CCGGNGGGCGGGCGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 144650

QY  1166  aatacgtgagctctaccaccagcgtgcgcgc 1197
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 144649 GGGCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 144618
```

Search completed: September 12, 2000, 13:35:21  
Job time: 12362 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2000, 10:09:19 ; Search time 52.14 Seconds  
(without alignments)  
5038.965 Million cell updates/sec

Title: US-09-135-238B-1  
Perfect score: 1910  
Sequence: 1 aaaggagtaagcagcgtgtc.....ttactctgtctccatcttt 1910

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 243080 seqs, 68777915 residues

Total number of hits satisfying chosen parameters: 486160

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/5C\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/5D\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/6\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq:\*  
7: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	56.2	2.9	7218	1	US-08-232-463-14
2	43.4	2.3	7218	1	US-08-232-463-14
3	38.6	2.0	1512	5	US-08-476-509B-1
4	38.6	2.0	1626	5	US-08-348-518C-1
5	38.2	2.0	1593	3	US-08-524-828-2
6	38.2	2.0	1593	3	US-08-975-114A-2
7	38.2	2.0	2247	3	US-08-524-828-1
8	38.2	2.0	2247	3	US-08-975-114A-1
9	37.6	2.0	5467	1	US-07-745-206A-12
10	37.6	2.0	5467	2	US-08-311-363-12
11	37.6	2.0	3089	3	US-08-472-934-5
12	37.6	2.0	3089	3	US-08-323-460A-5
13	37.6	2.0	3089	4	US-08-461-146C-5
14	37.6	2.0	3089	4	US-08-461-145C-5
15	36.8	1.9	4258	5	US-07-765-830A-5
16	36.8	1.9	3841	5	US-08-888-077A-33
17	36.4	1.9	1929	6	PCT-US93-00031-18
18	36.4	1.9	1929	6	PCT-US93-00031-20
19	36.4	1.9	1941	6	PCT-US93-00031-10
20	36.4	1.9	1941	6	PCT-US93-00031-22
21	36.4	1.9	2205	6	PCT-US93-00031-12
22	36.4	1.9	2208	6	PCT-US93-00031-14
23	36.4	1.9	2217	6	PCT-US93-00031-8
24	36.4	1.9	2220	6	PCT-US93-00031-16
25	36.4	1.9	2813	3	US-08-344-155C-99
26	36.2	1.9	6611	1	US-08-402-282-2

27	36.2	1.9	6611	1	US-08-508-004-2	Sequence 2, Appli
28	36.2	1.9	6611	1	US-08-402-066-2	Sequence 2, Appli
29	36.2	1.9	6611	1	US-08-402-068-2	Sequence 2, Appli
30	35.8	1.9	429	2	US-08-463-128-3	Sequence 3, Appli
31	35.8	1.9	429	2	US-08-463-298-3	Sequence 3, Appli
32	35.8	1.9	429	3	US-08-436-339A-3	Sequence 3, Appli
33	35.8	1.9	429	6	PCT-US94-10395-3	Sequence 3, Appli
34	35.8	1.9	2140	1	US-08-334-698-1	Sequence 1, Appli
35	35.8	1.9	2140	1	US-08-228-932-1	Sequence 1, Appli
36	35.8	1.9	2140	2	US-08-468-939-1	Sequence 1, Appli
37	35.8	1.9	2140	3	US-08-406-855A-1	Sequence 1, Appli
38	35.8	1.9	2140	4	US-08-722-190-1	Sequence 1, Appli
39	35.8	1.9	2140	5	US-08-244-354-1	Sequence 1, Appli
40	35.6	1.9	4695	6	PCT-US95-04203-1	Sequence 1, Appli
41	35.6	1.9	4695	3	US-08-231-193A-57	Sequence 57, Appl
42	35.6	1.9	4695	4	US-08-486-273A-57	Sequence 57, Appl
43	35.2	1.8	8438	1	US-07-945-283-1	Sequence 1, Appli
44	34.6	1.8	32207	3	US-08-770-379-20	Sequence 20, Appli
45	34.4	1.8	1894	5	US-08-749-816-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1  
US-08-232-463-14  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMM  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: pTZgpt-Fls  
US-08-232-463-14

```
Query Match          2.9%; Score 56.2; DB 1; Length 7210;
; Sequence Characteristics:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgt-F1s
US-08-232-463-14

Best Local Similarity 3.4%; Pred. No. 1.7e-06;
Matches 13; Conservative 223; Mismatches 151; Indels 0; Gaps 0;

QY 1051 gctcagcagcagggagggcccccgtcccgccgcggagcgcgctgcccgcgcgcgcg 1110
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1042 GCTCAGTGTGAGGAGTTCGATATATATATATATATATATATATATATATATATAT 1101
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1111 ctgcaggtgtcgaatctccctggctccatgcccacatctctgaagaccagctgtgaatac 1170
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1102 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1161
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1171 gtgagcctctaccaccgcctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1230
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1162 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1221
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1231 aatgtctcgtcgaactcccccagctatcccccacccagcgcgcgcgcgcgcgcgcgc 1290
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1222 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1281
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1291 aaggagctcactatctgtcgtgatcccaatcctgcttctcagagccctca 1350
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1282 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1341
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1351 tcactccatcccatctcgactcccatcccatcccatcccatcccatcccatcccatcc 1410
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1342 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1401
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1411 ctgccccaggtcgtctgcacacct 1437
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1402 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1428
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

## RESULT 2

```
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232.463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935.313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
```

```
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgt-F1s
US-08-232-463-14

Query Match          2.3%; Score 43.4; DB 1; Length 7218;
Best Local Similarity 3.0%; Pred. No. 0.01;
Matches 11; Conservative 207; Mismatches 153; Indels 0; Gaps 0;

QY 136 agaagtaagtagaggggagctgctgctgctgctgctgctgctgctgctgctgctgctga 195
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1415 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1356
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 196 aatgcattgagggatattctgtcgcggagatgctgctgctgctgctgctgctgctgct 255
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1355 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1296
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 256 ggtatccaccaccactcctcaagcgagatacaagcgcgagatacaagcgcgagcttctgaagcaata 315
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1295 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1236
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 316 cccagcagaaatctgttctcagtgaggttaacacagcgtcagcagaaagtgcagcagcag 375
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1235 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1176
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 376 ctatgcctgcggagcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 435
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1175 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1116
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 436 tgtccacagtgaatacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 495
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1115 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1056
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 496 gtttcattgc 506
| | | | |
Db 1055 CCTCAGCCTGC 1045
```

## RESULT 3

```
US-08-476-509B-1/c
; Sequence 1, Application US/08476509B
; Patent No. 6034212
; GENERAL INFORMATION:
; APPLICANT: SUDOL, MARIUS
; APPLICANT: PEER, BORR
; APPLICANT: HENRY, CHEN
; TITLE OF INVENTION: A SH2 DOMAIN ASSOCIATED PROTEIN, A
; TITLE OF INVENTION: SIGNALING DOMAIN THEREIN, NUCLEIC ACIDS ENCODING THE
; TITLE OF INVENTION: PROTEIN AND THE DOMAIN, AND DIAGNOSTIC AND THERAPEUTICUSES
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/476,509B
; FILING DATE: 01-DEC-1994
; CLASSIFICATION:
```

```

TELPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1626 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: Yap
FEATURE:
NAME/KEY: CDS
LOCATION: 66..1410
US-08-348-518C-1

```

	Query Match	2.0%;	Score 38.6;	DB 5;	Length 1626;
	Best Local Similarity	51.4%;	Pred. No. 0.11;	Mismatches 84;	Indels 0; Gaps 0;
	Matches 89;	Conservative	0;		
y	932	cgcgactggcgtgaagatgcgcccttgagagactcccagaggccccgcgggtgcgccq	991		
b	248	cttcggagtcgccccgcgcacatggacgatctgggtgcccccgcggaggggccccgcggccgg	189		
y	992	cqacgcgcctcccaaaacaatcatcacgcgctctgccgcgcgcctctggagcgcgacq	1051		
b	188	CGGCTGC CGGGCGCCTCCCGAGCTCCCGACCCGCCGCCGGGSGCTCGGGGCCCGCTG	129		
y	1052	ctfcgaggcacaggggagggcccccgcttccgcggccccggagcgcgttgtcccccc	1104		
	128	cmecmcegcacggcggtgcgcctgcctgcggcgctgctgcggcctgtagcctgccc	76		

RESULT 5  
US-08-524-828-2/c  
; Sequence 2, Application US/08524828  
; Patent No. 5874271  
; GENERAL INFORMATION:  
; APPLICANT: Atsushi NISHIKAWA et al.  
; TITLE OF INVENTION: HUMAN GLYCOSYLTRANSFERASE GENE,  
; TITLE OF INVENTION: COMPOUNDS AND METHOD FOR INHIBITING CANCEROUS METASTASIS  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wenderoth, Lind & Ponack  
; STREET: 805 Fifteenth Street, N.W., #700  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: Wordperfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/524,828  
; FILING DATE: September 7, 1995  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/107,173  
; FILING DATE: August 17, 1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: James H. Meadows  
; REGISTRATION NUMBER: 33,965  
; REFERENCE/DOCKET NUMBER: JTF/001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-371-8850  
; TELEFAX:  
; TELEX:

ATTORNEY/AGENT INFORMATION:  
NAME: JACKSON ESQ., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-101 CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201 487-5800  
TELEFAX: 201 343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1512 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 66..1409  
US-08-476-509B-1

	Query Match	2.0%;	Score 38.6;	DB 5;	Length 1512;
	Best Local Similarity	51.4%;	Pred. No. 0.11;	Mismatches 89;	Conservative 84; Indels 0; Gaps 0;
QY	932	cgcgactggcgctgaggatgcgcgccctagagcgtcccccagagacccccgcgagtgcgcg	991		
Dd	248	CTCGGAGTCCGCCCGGCACATGGACGATCTGGTCGCCCGCAGAGGGGCCCGGCCGG	189		
QY	992	cgaaccgcgcctcccaaaacaacatcatcacgacctgcgcgcgcgcgcctcttggagcgacg	1051		
Dd	188	CGGTGCGCGGGCGCCTCCC GGAGCTCCCGACCCCGCGCCGGGGGTGCGGGGCGCGCTG	129		
QY	1052	ctgcaggcacaggaggagcccccogettccccggccccggagcgcogttgtccccc	1104		
b)	128	CTCGGGSCCGGGSGCTGCGCGCTCGCGCGGCTGCTCGGGTAGAGCTGCC	76		

```

RESULT      4
US-08-348-518C-1/c
; Sequence 1, Application US/08348518C
; Patent No. 6022740
; GENERAL INFORMATION:
; APPLICANT: SUDOL, MARIUS
; APPLICANT: PEER, BORK
; APPLICANT: HENRY, CHEN
; TITLE OF INVENTION: A SH2 DOMAIN ASSOCIATED PROTEIN, A
; TITLE OF INVENTION: SIGNALLING DOMAIN THEREIN, NUCLEIC ACIDS ENCODING THE
; TITLE OF INVENTION: PROTEIN AND THE DOMAIN, AND DIAGNOSTIC AND THERAPEUTIC USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/348,518C
; FILING DATE: 01-DEC-1994
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-101
; TELECOMMUNICATION INFORMATION:

```



```
Db 607 GTGGCCGCGGCGGAGGCGGTGGCCCTCAGGCTTCTCTCCGGCGGTCCCGGGG 548
QY 1035 gcgcctctgagcgagcgtcagggcacagggagggcccccgttccccggcccgagcg 1093
Db 547 GCGGCTCTCCAGCATCTTGGTGGCGGTTTGAAGCAGACCGCGCGCTTGGTGCGC 489

RESULT 8
US-08-975-114A-1/c
; Sequence 1, Application US/08975114A
; Patent No. 5876714
; GENERAL INFORMATION:
; APPLICANT: Alesushi NISHIKAWA et al.
; TITLE OF INVENTION: HUMAN GLYCOSYLTRANSFERASE GENE,
; TITLE OF INVENTION: COMPOUNDS AND METHOD FOR INHIBITING CANCEROUS METASTASIS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., Suite 800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/975-114A
; FILING DATE: No. 5876714ember 20, 1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/107,173
; FILING DATE: August 17, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee Cheng
; REGISTRATION NUMBER: 40,949
; REFERENCE/DOCKET NUMBER: 1-F3439DIV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2247 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; US-08-975-114A-1

Query Match 2.0%; Score 38.2; DB 3; Length 2247;
Best Local Similarity 50.8%; Pred. No. 0.18;
Matches 91; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 915 ccctctccagggcgccgcgactgagcgatgagcgccctggagagctccaga 974
Db 667 CCACATTGCGCGGCGCCTCGGCCCGCTCCCGGCGCTCAGAGGTACCGG 608
QY 975 ggcgcgcggtgcgcgcgacgcgtcccaaaacatctacagcgctgccgcgcgc 1034
Db 607 GTGGCCGCGGCGGAGAGCGCTTGGCCCTCAGGCTTCTCTCGGCCGTCCCGGG 548
QY 1035 gcgcctggagcgagcgtcagggcacagggagggcccccgttccccggcccgagcg 1093
Db 547 GCGGCTCTCCAGCATCTTGGTGGCGGTTTGAAGCAGACCGCGCGCTTGGTGCGC 489

RESULT 9
US-07-745-206A-12
; Sequence 12, Application US/07745206A
```

```
; Patent No. 5429921
; GENERAL INFORMATION:
; APPLICANT: Harpold, Michael
; APPLICANT: Ellis, Steven
; APPLICANT: Williams, Mark
; APPLICANT: McCue, Ann
; APPLICANT: Feldman, Daniel
; TITLE OF INVENTION: Human Calcium Channel Compositions and
; TITLE OF INVENTION: Methods
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fitch, Even, Tabin & Flannery
; STREET: 135 S. LaSalle
; CITY: Chicago
; STATE: Illinois
; COUNTRY: U.S.A.
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/745,206A
; FILING DATE: 19910815
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Feder, Scott B
; REFERENCE/DOCKET NUMBER: 51504
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-372-7842
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5467 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Join(144...3164, 3168..3245, 3249..3386, 3390
; LOCATION: ..3392, 3396..3488, 3495..3539, 3543..3581, 3585
; LOCATION: ..3587, 3591..3626, 3630..3689, 3693..3737, 3744
; LOCATION: ..3746, 3750..4823, 4827..4841, 4845..5006, 5010
; LOCATION: ..5096, 5100..5306, 5310..5366, 5370..5465)
; US-07-745-206A-12

Query Match 2.0%; Score 37.6; DB 1; Length 5467;
Best Local Similarity 51.8%; Pred. No. 0.46;
Matches 85; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 961 ggagagctccagagggcccggggtcgcggagcgctcccaaaacatctacag 1020
Db 2765 GGAGAGCGGGAGCGCGGTGCGCGGAGGCGCGCGCGCGCGCGCGCGCGCGCA 2824
QY 1021 gccttgcgcgcgcgcgtctctgagcgacgctgcagcgagggagggcccggtccc 1080
Db 2825 GGAGCGCGGGGCGCGCGCGGAGCGCGGACGAGCGCGCGCGCGCGCGCGG 2884
QY 1081 ggcgcgcgagcgccgttcccccccccgccgcgtgcagggtgtctga 1124
Db 2885 GCGCGCGGCGCACCGCGCGCGCTCCCGGAGGAGCGGCGCGCA 2928

RESULT 10
US-08-311-363-12
; Sequence 12, Application US/08311363
; Patent No. 5876958
; GENERAL INFORMATION:
; APPLICANT: Harpold, Michael
; APPLICANT: Ellis, Steven
; APPLICANT: Williams, Mark
```









Tue Sep 12 15:26:55 2000

Db 1382 CTTGGAGCTGCAGAGCTGAGCGCACAGGGGAACCTCCATTTCAGCGAGCGCGCCCTGG 1323  
Qy 1074 cgtcccgcccgagcgccgttgcggcccgcccgctgcagggtgtctgaatctccctg 1133  
Db 1322 CCTGCCGAGCATCGGTGCGCCTTGGTGCCCTTGTGACCCACTAGGTCTCATCGCACTC 1263  
Qy 1134 gctcca 1139  
Db 1262 ACTCTA 1257

Search completed: September 12, 2000, 11:49:07  
Job time: 5988 sec



\*\*\*\*\*

\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

```

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Aug 24 10:25:10 2000; MasPar time 9.27 Seconds
Tabular output not generated. 644.932 Million cell updates/sec

```

```
>US-09-135-238B-2
Title:
Description:
Perfect Score: 2830
Sequence: 1 MDRWLPLYFLPVSGALRIIL.....HOPAAAMMEDSDSDYINPVA 390
```

Scoring table: PAM 150  
Gap 11

Searched: 152433 seqs, 15329240 residues

Post-processing: Minimum Match 0% Listing first 45 summaries

```
Database:
a-issued
1:5A_COMB 2:5B_COMB 3:6_COMB 4:PCT_COMB 5:backfiles1
```

Statistics: Mean 34.052; Variance 186.097; scale 0.183

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query			DB	ID	Description	Pred. No.
	Score	Match	Length				
1	254	9.0	771	3	US-08-434-	Sequence 8, Applicatio	3.28e-09
2	247	8.7	746	3	US-08-434-	Sequence 4, Applicatio	1.02e-08
3	245	8.7	757	3	US-08-434-	Sequence 6, Applicatio	1.42e-08
4	246	8.7	769	3	US-08-434-	Sequence 10, Applicati	1.20e-08
5	223	7.9	624	2	US-08-642-	Sequence 22, Applicati	4.91e-07
6	223	7.9	773	3	US-08-434-	Sequence 2, Applicatio	4.91e-07
7	136	4.8	560	1	US-08-336-	Sequence 22, Applicati	3.32e-01
8	129	4.6	201	3	US-08-955-	Sequence 2, Applicatio	9.18e-01
9	125	4.4	46	3	US-08-955-	Sequence 10, Applicati	1.63e+00
10	123	4.3	195	3	US-08-955-	Sequence 4, Applicatio	2.17e+00
11	122	4.3	303	1	US-08-198-	Sequence 2, Applicatio	2.50e+00
12	122	4.3	303	1	US-08-323-	Sequence 2, Applicatio	2.50e+00
13	122	4.3	303	4	PCT-US95-0	Sequence 2, Applicatio	2.50e+00
14	122	4.3	303	4	PCT-US93-0	Sequence 2, Applicatio	2.50e+00
15	118	4.2	319	1	US-08-597-	Sequence 22, Applicati	4.41e+00
16	112	4.0	57	3	US-08-955-	Sequence 5, Applicatio	1.02e-01
17	112	4.0	60	3	US-08-955-	Sequence 6, Applicatio	1.02e-01
18	109	3.9	114	1	US-08-211-	Sequence 23, Applicati	1.55e+01
19	109	3.9	114	4	PCT-US93-0	Sequence 23, Applicati	1.55e+01
20	109	3.9	114	1	US-08-111-	Sequence 23, Applicati	1.55e+01
21	110	3.9	134	2	US-08-822-	Sequence 10, Applicati	1.55e+01
22	108	3.8	76	3	US-08-554-	Sequence 21, Applicati	1.78e+01
23	108	3.8	87	3	US-08-554-	Sequence 18, Applicati	1.78e+01

```
CC      TELEX: 67-3510
CC      TELETYPE: SEQUENCE LISTING
CC      INFORMATION FOR SEQ ID NO: 8:
CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH: 771 amino acids
CC      TYPE: amino acid
CC      STRANDEDNESS: single
CC      TOPOLOGY: linear
CC      DESCRIPTION: Mouse Polyimmunoglobulin Receptor
SQ      SEQUENCE 771 AA; 85097 MW; 3010231 CN;

Query Match      9.0%; Score 254; DB 3; Length 771;
Best Local Similarity 36.8%; Pred. No. 3.28e-09;
Matches 43; Conservative 26; Mismatches 42; Indels 6; Gaps 5;

Db 13 FSGVSTKPIFGQEVSSITGDSVSIYCPDTSVNRHTRKYWCRO--GASGMCTLLISN 71
Qy 10 FLFVSGALRIILPEVKVEGLGSGVTIKC--P-LP-EMHVRILYLCREMAGSGTCGVIVSTT 65
Db 72 GYLSKEYSGRANLINFENNFTVINIEQLTQDDTSYKQGLG-TSNRGLSFDVSLV 127
Qy 66 NFIKAEYGRVTLKQYPRKNLFLVEVTQLTESDSGYVACGAGMNTDRGKTQKVTLVN 122

RESULT 2
ID US-08-434-000A-4 STANDARD; PRT; 746 AA.
XX
AC xxxxxx
XX
DT
DT
XX
XX
DE
XX
XX
Sequence 4, Application US/08434000A
XX
Sequence 4, Application US/08434000A
CC      Patent No. 6046037
CC      GENERAL INFORMATION:
CC      APPLICANT: ANDREW C. HIATT, JULIAN
CC      APPLICANT: K.-C. MA, THOMAS LEHNER
CC      TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
CC      TITLE OF INVENTION: PROTEINS IN PLANTS AND THEIR USES
CC      NUMBER OF SEQUENCES: 19
CC      CORRESPONDENCE ADDRESS:
CC      ADDRESSEE: Lyon & Lyon
CC      STREET: 633 West Fifth Street
CC      STREET: Suite 4700
CC      CITY: Los Angeles
CC      STATE: California
CC      COUNTRY: U.S.A.
CC      ZIP: 90071
CC      COMPUTER READABLE FORM:
CC      MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
CC      MEDIUM TYPE: storage
CC      COMPUTER: IBM Compatible
CC      OPERATING SYSTEM: IBM P.C. DOS 5.0
CC      SOFTWARE: Word Perfect 5.1
CC      CURRENT APPLICATION DATA:
CC      APPLICATION NUMBER: US/08/434,000A
CC      FILING DATE:
CC      CLASSIFICATION: 435
CC      PRIOR APPLICATION DATA: including application
CC      PRIOR APPLICATION DATA: described below: 1
CC      PRIOR APPLICATION DATA: described below:
CC      APPLICATION NUMBER: 08/367,395
CC      FILING DATE: 12/30/94
CC      ATTORNEY/AGENT INFORMATION:
CC      NAME: Guise, Jeffrey W.
CC      REGISTRATION NUMBER: 34,613
CC      REFERENCE/DOCKET NUMBER: 212/127
CC      TELECOMMUNICATION INFORMATION:
CC      TELEPHONE: (619) 552-8400
CC      TELEFAX: (619) 552-0159
CC      TELEX: 67-3510
CC      TELETYPE: SEQUENCE LISTING
CC      SEQUENCE CHARACTERISTICS:

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 746 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
DESCRIPTION: Human Polyimmunoglobulin Receptor
SQ SEQUENCE 746 AA; 81378 MW; 2795658 CN;

Query Match      8.7%; Score 247; DB 3; Length 746;
Best Local Similarity 37.7%; Pred. No. 1.02e-08;
Matches 40; Conservative 27; Mismatches 32; Indels 7; Gaps 5;

Db 7 PE-EVNSVGNVSITCYPTSVNRHTRKYWCRO--GARGGCITLISSEGVSSKYAGRA 64
Qy 21 PEVKVEGLGSGVTIKC---PLP-EMHVRILYLCREMAGSGTCGVIVSTTNIKAEYGRV 76
Db 65 NLNFPENSTFVNNIAQLSQDSDGRYKCGLGINS-RGLSFDVSLV 109
Qy 77 TLQYPRKNLFLVEVTQLTESDSGYVACGAGMNTDRGKTQKVTLVN 122

RESULT 3
ID US-08-434-000A-6 STANDARD; PRT; 757 AA.
XX
AC xxxxxx
XX
DT
DT
XX
XX
DE
XX
XX
Sequence 6, Application US/08434000A
XX
Sequence 6, Application US/08434000A
CC      Patent No. 6046037
CC      GENERAL INFORMATION:
CC      APPLICANT: ANDREW C. HIATT, JULIAN
CC      APPLICANT: K.-C. MA, THOMAS LEHNER
CC      TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
CC      TITLE OF INVENTION: PROTEINS IN PLANTS AND THEIR USES
CC      NUMBER OF SEQUENCES: 19
CC      CORRESPONDENCE ADDRESS:
CC      ADDRESSEE: Lyon & Lyon
CC      STREET: 633 West Fifth Street
CC      STREET: Suite 4700
CC      CITY: Los Angeles
CC      STATE: California
CC      COUNTRY: U.S.A.
CC      ZIP: 90071
CC      COMPUTER READABLE FORM:
CC      MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
CC      MEDIUM TYPE: storage
CC      COMPUTER: IBM Compatible
CC      OPERATING SYSTEM: IBM P.C. DOS 5.0
CC      SOFTWARE: Word Perfect 5.1
CC      CURRENT APPLICATION DATA:
CC      APPLICATION NUMBER: US/08/434,000A
CC      FILING DATE:
CC      CLASSIFICATION: 435
CC      PRIOR APPLICATION DATA: including application
CC      PRIOR APPLICATION DATA: described below: 1
CC      PRIOR APPLICATION DATA: described below:
CC      APPLICATION NUMBER: 08/367,395
CC      FILING DATE: 12/30/94
CC      ATTORNEY/AGENT INFORMATION:
CC      NAME: Guise, Jeffrey W.
CC      REGISTRATION NUMBER: 34,613
CC      REFERENCE/DOCKET NUMBER: 212/127
CC      TELECOMMUNICATION INFORMATION:
CC      TELEPHONE: (619) 552-8400
CC      TELEFAX: (619) 552-0159
CC      TELEX: 67-3510
CC      TELETYPE: SEQUENCE LISTING
CC      INFORMATION FOR SEQ ID NO: 6:
CC      SEQUENCE CHARACTERISTICS:
```

CC LENGTH: 757 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC TOPOLOGY: DESCRIPTION: Bovine Polymunoglobulin Receptor  
SQ SEQUENCE 757 AA; 82450 MW; 2942196 CN;

Query Match 8.7%; Score 245; DB 3; Length 757;  
Best Local Similarity 36.7%; Pred. No. 1.42e-08;  
Matches 40; Conservative 25; Mismatches 37; Indels 7; Gaps 5;

Db 25 PF-EVSVSVEGRSVSKYPTSPVNRHTRKYWCRO-GAQRCTLTLSSEGYVSDDYVGRA 82  
QY 21 PEVKVEGELGGSVTKC---PLP-EMHVRVYLCREMAGSGTCGTVSTTTFIKAEYKGRV 76  
Db 83 NLTNPESGTFVVDISHLTHDKSGRYKCGLGISS-RGLNFDVLSLEVSQD 130  
QY 77 TLKQTPRKNLFLVEVTQLTESDSGVYACGAGMNTDRGKTQKVTNLVHSE 125

RESULT 4  
ID US-08-434-000A-10 STANDARD; PRT; 769 AA.  
XX  
AC xxxxxx  
XX  
DT  
XX  
DE  
XX  
Sequence 10, Application US/08434000A  
Sequence 10, Application US/08434000A  
Patent No. 6045037  
GENERAL INFORMATION:  
APPLICANT: ANDREW C. HIATT, JULIAN  
APPLICANT: K.-C. MA, THOMAS LEHNER  
TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION  
TITLE OF INVENTION: PROTEINS IN PLANTS AND THEIR USES  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LYON & LYON  
STREET: 633 West Fifth Street  
CITY: Suite 4700  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/434,000A  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
PRIOR APPLICATION DATA: including application  
PRIOR APPLICATION DATA: described below: 1  
APPLICATION NUMBER: 08/367,395  
FILING DATE: 12/30/94  
ATTORNEY/AGENT INFORMATION:  
NAME: Guise, Jeffrey W.  
REGISTRATION NUMBER: 34,613  
REFERENCE/DOCKET NUMBER: 212/127  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 552-8400  
TELEFAX: (619) 552-0159  
TELEX: 67-3510  
TELEX: SEQUENCE LISTING  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 769 amino acids  
TYPE: amino acid

CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC TOPOLOGY: DESCRIPTION: Rat Polymunoglobulin Receptor  
SQ SEQUENCE 769 AA; 84798 MW; 2960864 CN;

Query Match 8.7%; Score 246; DB 3; Length 769;  
Best Local Similarity 35.0%; Pred. No. 1.20e-08;  
Matches 41; Conservative 30; Mismatches 40; Indels 6; Gaps 5;

Db 13 FSGYSTOSPFEGQDYSVSIENSVSITCYPPDPSVNRHTRKYWCRO-GANGYCATLISSN 71  
QY 10 FLPSVSGALRLPEVKVVEGELGGSVTKC--P-LP-EMHVRVYLCREMAGSGTCGTVWST 65  
Db 72 GYLSKEYSGRASLINEFENSTFVINIAHLTQEDTGSYKGLG-TTNRGLFFDVSLEV 127  
QY 66 NFIKAEYKGRVTLKQYPRKNLFLVEVTQLTESDSGVYACGAGMNTDRGKTQKVTNLV 122

RESULT 5  
ID US-08-642-406A-22 STANDARD; PRT; 624 AA.  
XX  
AC xxxxxx  
XX  
DT  
XX  
DE  
XX  
Sequence 22, Application US/08642406A  
Sequence 22, Application US/08642406A  
Patent No. 5959177  
GENERAL INFORMATION:  
APPLICANT: Hein, Mich B.  
APPLICANT: Hiatt, Andrew C.  
APPLICANT: Ma, Julian K.C.  
TITLE OF INVENTION: TRANSGENIC PLANTS EXPRESSING ASSEMBLED  
TITLE OF INVENTION: SECRETORY ANTIBODIES  
NUMBER OF SEQUENCES: 26  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: THE SCRIPPS RESEARCH INSTITUTE  
STREET: 10666 No. 5959177th Torrey Pines Road, TPC-8  
CITY: La Jolla  
STATE: California  
COUNTRY: US  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/642,406A  
FILING DATE: 03-MAY-1996  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/591,823  
FILING DATE: 02-OCT-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/427,765  
FILING DATE: 27-OCT-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Logan, April C.  
REGISTRATION NUMBER: 33,950  
REFERENCE/DOCKET NUMBER: 184.2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 554-2937  
TELEFAX: (619) 554-6312  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 624 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE 624 AA; 68416 MW; 2044676 CN;





AC xxxxxx  
XX  
DT  
XX  
DE Sequence 4, Application US/08955937A  
CC  
CC Sequence 4, Application US/08955937A  
CC Patent No. 6020161  
CC GENERAL INFORMATION:  
CC APPLICANT: WU, SHUIJIAN  
CC APPLICANT: SWEET, RAYMOND  
CC APPLICANT: TRUNEH, ALEMSEGED  
CC TITLE OF INVENTION: FIGR-1, A MEMBER OF IMMUNOGLOBULIN  
CC NUMBER OF SEQUENCES: 14  
CC CORRESPONDENCE ADDRESS:  
CC STREET: P.O. BOX 980  
CC CITY: VALLEY FORGE  
CC STATE: PA  
CC COUNTRY: USA  
CC ZIP: 19482  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Diskette  
CC OPERATING SYSTEM: DOS  
CC SOFTWARE: FASTSEQ for Windows Version 2.0  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/955,937A  
CC FILING DATE: 17-OCT-1997  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 60/056,152  
CC FILING DATE: 19-AUG-1997  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: PRESTIA, PAUL F  
CC REGISTRATION NUMBER: 23,031  
CC REFERENCE/DOCKET NUMBER: GH-70228  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 610-407-0700  
CC TELEFAX: 610-407-0701  
CC TELEX: 846169  
CC INFORMATION FOR SEQ ID NO: 4:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 195 amino acids  
CC TYPE: amino acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 195 AA; 22248 MW; 210677 CN;  
Query Match 4.3%; Score 123; DB 3; Length 195;  
Best Local Similarity 29.5%; Pred. No. 2.17e+00;  
Matches 31; Conservative 18; Mismatches 52; Indels 4; Gaps 4;  
Db 2 WLPLALLLSGCFSGIOPESVRAPEQSLTVQCHYKQGWETIYKWCGRGYR-WDTCKI 60  
Qy 4 WLWP-LYFLPVGALRIILPEVKVEGLGGSVTKICPLPEM-HVRI-YLCREMAGSGTCGT 60  
Db 61 LIETRGSQGEKSDRYSINKDKRDTFTVTMEGLRDRDDADVTWG 105  
Qy 61 VYSTNTFKAEYKGRVTLKQYPRKNLFLVETLTSDSGVIYAG 105  
RESULT 11  
ID US-08-198-094-2 STANDARD; PRT; 303 AA.  
XX  
AC xxxxxx  
XX  
DT  
XX  
DE Sequence 2, Application US/08198094

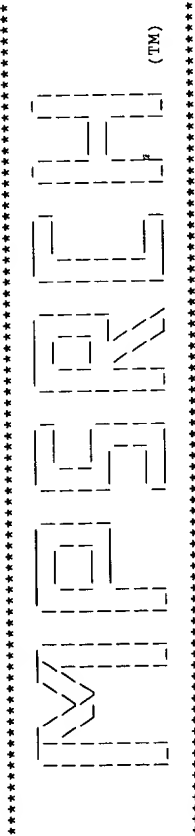
CC  
XX Sequence 2, Application US/08198094  
CC Patent No. 5741696  
CC GENERAL INFORMATION:  
CC APPLICANT: Cochran Ph.D., Mark D  
CC TITLE OF INVENTION: Recombinant Equine Herpesvirus  
CC NUMBER OF SEQUENCES: 77  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: John P. White  
CC STREET: 30 Rockefeller Plaza  
CC CITY: New York  
CC STATE: New York  
CC COUNTRY: USA  
CC ZIP: 10112  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/198,094  
CC FILING DATE: February 17, 1994  
CC CLASSIFICATION: 435  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: White, John P  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (212)977-9550  
CC TELEFAX: (212)664-0525  
CC TELEX: 422523  
CC INFORMATION FOR SEQ ID NO: 2:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 303 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 303 AA; 33465 MW; 451890 CN;  
Query Match 4.3%; Score 122; DB 1; Length 303;  
Best Local Similarity 32.4%; Pred. No. 2.50e+00;  
Matches 22; Conservative 19; Mismatches 23; Indels 4; Gaps 4;  
Db 224 PARLEIPAPGRPLPRRP-SEGGM-RAPRRSRAPAAASTAAATPPRGCDPAPAAAR 281  
Qy 290 AVRMALRESQRPGRSPRPSQNNIYSACPRRAGADAA-GTGEAPVG-PGAPLPPAPL 347  
Db 282 RAGDVTWM 289  
Qy 348 QVSESPWL 355  
RESULT 12  
ID US-08-323-531-2 STANDARD; PRT; 303 AA.  
XX  
AC xxxxxx  
XX  
DT  
XX  
DE Sequence 2, Application US/08323531  
XX  
CC  
CC Sequence 2, Application US/08323531  
CC Patent No. 5731188  
CC GENERAL INFORMATION:  
CC APPLICANT: Cochran Ph.D., Mark D  
CC TITLE OF INVENTION: Recombinant Equine Herpesvirus  
CC NUMBER OF SEQUENCES: 71  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: John P. White  
CC STREET: 30 Rockefeller Plaza  
CC CITY: New York  
CC STATE: New York  
CC COUNTRY: USA  
CC ZIP: 10112  
CC COMPUTER READABLE FORM: disk  
CC MEDIUM TYPE: Floppy disk



CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/323,531  
CC FILING DATE:  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: US/07/936,784  
CC FILING DATE:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: White, John P  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (212)977-9550  
CC TELEFAX: (212)664-0525  
CC TELEX: 422523  
CC INFORMATION FOR SEQ ID NO: 2:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 303 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 303 AA; 33465 MW; 451890 CN;  
CC  
CC Query Match 4.3%; Score 122; DB 1; Length 303;  
CC Best Local Similarity 32.4%; Pred. No. 2.50e+00;  
CC Matches 22; Conservative 19; Mismatches 23; Indels 4; Gaps 4;  
CC  
Db 224 PARLEIPAPGRPLPRRP-SEGGM-RAPRRSRAPAAARSTAAATPPRPGDPAPAAAR 281  
Qy 290 AVRMRALSSORPGSRPRSRNNIYACPRRAGADAA-GTGEAPVPG-PGAPLPAPL 347  
Db 282 RAGDVTWM 289  
Qy 348 QVSESPWL 355  
CC  
CC RESULT 13  
CC ID PCT-US95-02087-2 STANDARD: PRT: 303 AA.  
CC XX xxxxxx  
CC  
CC DE Sequence 2, Application PC/TUS9502087  
CC  
CC Sequence 2, Application PC/TUS9502087  
CC GENERAL INFORMATION:  
CC APPLICANT: Cochran, Mark D  
CC TITLE OF INVENTION: Recombinant Equine Herpesviruses  
CC NUMBER OF SEQUENCES: 77  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: John P. White  
CC STREET: 1185 Avenue of the Americas  
CC CITY: New York  
CC STATE: New York  
CC COUNTRY: USA  
CC ZIP: 10036  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: PCT/US95/02087  
CC FILING DATE:  
CC CLASSIFICATION:  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER: 08/198,094  
CC FILING DATE: February 17, 1994  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: White, John P  
CC REGISTRATION NUMBER: 28,678

CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (212)278-0400  
CC TELEFAX: (212)391-0525  
CC TELEX: 422523  
CC INFORMATION FOR SEQ ID NO: 2:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 303 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 303 AA; 33465 MW; 451890 CN;  
CC  
CC Query Match 4.3%; Score 122; DB 4; Length 303;  
CC Best Local Similarity 32.4%; Pred. No. 2.50e+00;  
CC Matches 22; Conservative 19; Mismatches 23; Indels 4; Gaps 4;  
CC  
Db 224 PARLEIPAPGRPLPRRP-SEGGM-RAPRRSRAPAAARSTAAATPPRPGDPAPAAAR 281  
Qy 290 AVRMRALSSORPGSRPRSRNNIYACPRRAGADAA-GTGEAPVPG-PGAPLPAPL 347  
Db 282 RAGDVTWM 289  
Qy 348 QVSESPWL 355  
CC  
CC RESULT 14  
CC ID PCT-US93-07424-2 STANDARD: PRT: 303 AA.  
CC XX xxxxxx  
CC  
CC DE Sequence 2, Application PC/TUS9307424  
CC  
CC Sequence 2, Application PC/TUS9307424  
CC GENERAL INFORMATION:  
CC APPLICANT: Cochran Ph.D., Mark D  
CC TITLE OF INVENTION: Recombinant Equine Herpesvirus  
CC NUMBER OF SEQUENCES: 71  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: John P. White  
CC STREET: 30 Rockefeller Plaza  
CC CITY: New York  
CC STATE: New York  
CC COUNTRY: USA  
CC ZIP: 10112  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: PatentIn Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: PCT/US93/07424  
CC FILING DATE: 19930806  
CC CLASSIFICATION:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: White, John P  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (212)977-9550  
CC TELEFAX: (212)664-0525  
CC TELEX: 422523  
CC INFORMATION FOR SEQ ID NO: 2:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 303 amino acids  
CC TYPE: amino acid  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: protein  
CC SEQUENCE 303 AA; 33465 MW; 451890 CN;  
CC  
CC Query Match 4.3%; Score 122; DB 4; Length 303;  
CC Best Local Similarity 32.4%; Pred. No. 2.50e+00;  
CC Matches 22; Conservative 19; Mismatches 23; Indels 4; Gaps 4;  
CC





Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Aug 24 10:22:24 2000; MasPar time 13.89 Seconds  
870.398 Million cell updates/sec

Tabular output not generated.

Title: >US-09-135-238B-2  
Description: (1-390) from US09135238B.pep  
Perfect Score: 2830  
Sequence: 1 MDRWLWFLYFLVPVSGALRIL.....HQPAAEMEDSDSDYINVPA 390

Scoring table: PAM 150  
Gap 11

Searched: 85661 seqs, 30989116 residues

Post-processing: Minimum Match 0%

Database: swiss-prot38  
1:swissprot

Statistics: Mean 50.211; Variance 94.729; scale 0.530

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	254	9.0	771	1	P1GR_MOUSE POLYMERIC-IMMUNOGLOBUL	1.08e-25
2	252	8.9	764	1	P1GR_HUMAN POLYMERIC-IMMUNOGLOBUL	2.54e-25
3	246	8.7	769	1	P1GR_RAT POLYMERIC-IMMUNOGLOBUL	3.32e-24
4	244	8.6	757	1	P1GR_BOVIN POLYMERIC-IMMUNOGLOBUL	7.79e-24
5	223	7.9	773	1	P1GR_RABIT POLYMERIC-IMMUNOGLOBUL	5.43e-20
6	188	6.6	102	1	P1GR_PIG POLYMERIC-IMMUNOGLOBUL	8.21e-14
7	140	4.9	224	1	CM35_HUMAN CMRF35 ANTIGEN PRECURS	5.46e-06
8	126	4.5	3321	1	KENDRIN (K1AA0402).	6.49e-04
9	119	4.2	303	1	US02_HSVEX US1 PROTEIN	6.35e-03
10	118	4.2	319	1	A33_HUMAN CELL SURFACE A33 ANTIG	8.74e-03
11	115	4.1	116	1	STP2_RAT NUCLEAR TRANSITION PRO	2.25e-02
12	115	4.1	179	1	V559_HSVB HYPOTHETICAL GENE 59 P	1.20e-02
13	117	4.1	567	1	CYDC_BAGSU TRANSPORT ATP-BINDING	3.08e-02
14	114	4.0	117	1	STP2_MOUSE NUCLEAR TRANSITION PRO	3.08e-02
15	114	4.0	146	1	HY21_HUMAN IG HEAVY CHAIN V-II RE	4.20e-02
16	113	4.0	863	1	MCM4_HUMAN DNA REPLICATION LICENS	1.43e-01
17	109	3.9	502	1	VPHE_NPVAC POLYHEDRAL ENVELOPE PR	1.93e-01
18	108	3.8	252	1	P2Y_HUMAN P2Y PURINOCEPTOR 7 (P2	1.93e-01
19	108	3.8	352	1	VCOM_ADE40 MINOR CORE PROTEIN (PR	1.93e-01
20	108	3.8	358	1	YXK2_CAEBL HYPOTHETICAL 24.0 KDA	6.28e-01
21	104	3.7	208	1	CSP_PLABA CIRCUMSPOROZOITE PROTE	6.28e-01
22	104	3.7	347	1	VE2_HPV14 REGULATORY PROTEIN E2.	6.28e-01
23	104	3.7	483	1		

RESULT ID	P1GR_MOUSE	STANDARD;	PRT;	771 AA.
AC	070570;			
DT	15-FEB-2000 (Rel. 39, Created)			
DT	15-FEB-2000 (Rel. 39, Last sequence update)			
DT	15-FEB-2000 (Rel. 39, Last annotation update)			
DE	POLYMERIC-IMMUNOGLOBULIN RECEPTOR PRECURSOR (POLY-IG RECEPTOR) (P1GR)			
DE	[CONTAINS: SECRETORY COMPONENT].			
GN	P1GR.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6 X CBA; TISSUE=LIVER;			
RX	MEDLINE; 95138517.			
RA	Piskurich J.F., Blanchard M.H., Youngman K.R., France J.A.,			
RA	Kaetzel C.S.;			
RT	"Molecular cloning of the mouse polymeric Ig receptor. Functional			
RT	regions of the molecule are conserved among five mammalian species.";			
RL	J. Immunol. 154:1735-1747(1995).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=129/SVJ;			
RX	MEDLINE; 98072444.			
RA	Martin M.G., Gutierrez E.M., Lam J.T., Li T.W.H., Wang J.;			
RT	"Genomic cloning and structural analysis of the murine polymeric			
RT	receptor (p1gr) gene and promoter region.";			
RL	Gene 201:189-197(1997).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=129; TISSUE=LIVER;			
RA	de Groot N., Vollebregt E., Lee S.H., Verbeet M.P., de Boer H.A.;			
RT	"Molecular cloning and exon-intron organization of the gene encoding			
RT	the murine polymeric immunoglobulin receptor.";			
RL	Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.			
CC	-1- FUNCTION: THIS RECEPTOR BINDS POLYMERIC IGA AND IGM AT THE			
CC	BASOLATERAL SURFACE OF EPITHELIAL CELLS. THE COMPLEX IS THEN			
CC	TRANSPORTED ACROSS THE CELL TO BE SECRETED AT THE APICAL SURFACE.			
CC	DURING THIS PROCESS A CLEAVAGE OCCURS THAT SEPARATE THE			
CC	EXTRACELLULAR (KNOWN AS THE SECRETORY COMPONENT) FROM THE			
CC	TRANSMEMBRANE SEGMENT (BY SIMILARITY).			
CC	-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ALSO SECRETED.			
CC	-1- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE V-TYPE DOMAINS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			

24	105	3.7	575	1	SMF1_YEAST	TRANSPORTER PROTEIN SM	4.69e-01
25	104	3.7	1403	1	VG22_HSV11	HYPOTHETICAL GENE 22 P	6.28e-01
26	103	3.6	339	1	CSP_PLABE	CIRCUMSPOROZOITE PROTE	8.39e-01
27	102	3.6	354	1	CD68_HUMAN	MACROSIALIN PRECURSOR	1.12e+00
28	101	3.6	384	1	YKAP_ECOLI	HYPOTHETICAL 43.6 KDA	1.45e+00
29	102	3.6	876	1	AREA_EMEI	NITROGEN REGULATORY PR	1.12e+00
30	102	3.6	1523	1	SON_HUMAN	SON PROTEIN (SON3).	1.12e+00
31	98	3.5	178	1	TRSF_DROER	FEMALE-SPECIFIC TRANSF	3.46e+00
32	99	3.5	199	1	COAT_BYDVM	COAT PROTEIN.	2.62e+00
33	99	3.5	199	1	COAT_BYDV1	COAT PROTEIN.	2.62e+00
34	100	3.5	393	1	XYLH_ECOLI	XYLOSE TRANSPORT SYSTE	1.98e+00
35	100	3.5	463	1	ROK_HUMAN	HETEROGENEOUS NUCLEAR	2.62e+00
36	99	3.5	467	1	VE2_HPV24	REGULATORY PROTEIN E2.	3.46e+00
37	98	3.5	471	1	RUI7_XENLA	UI SMALL NUCLEAR RIBON	3.46e+00
38	98	3.5	778	1	FASK_MOUSE	FATTY ACID SYNTHASE (E	3.46e+00
39	98	3.5	838	1	MCM4_MOUSE	DNA REPLICATION LICENS	2.62e+00
40	98	3.5	862	1	MCM4_XENLA	DNA REPLICATION LICENS	2.62e+00
41	99	3.5	863	1	MML3_MYCTU	PUTATIVE MEMBRANE PROT	2.62e+00
42	99	3.5	944	1	CCAA_RABIT	VOLTAGE-DEPENDENT P/Q-	2.62e+00
43	99	3.5	2424	1	TRA3_ECOLI	TRANSPORASE FOR TRANSP	4.56e+00
44	97	3.4	988	1	HFC1_HUMAN	HOST CELL FACTOR C1 (H	4.56e+00
45	97	3.4	2035	1			

ALIGNMENTS

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: U06431; AAC67440.1; -  
 DR EMBL: U83434; AAC53585.1; -  
 DR EMBL: U83427; AAC53585.1; JOINED.  
 DR EMBL: U83428; AAC53585.1; JOINED.  
 DR EMBL: U83429; AAC53585.1; JOINED.  
 DR EMBL: U83430; AAC53585.1; JOINED.  
 DR EMBL: U83431; AAC53585.1; JOINED.  
 DR EMBL: U83432; AAC53585.1; JOINED.  
 DR EMBL: U83433; AAC53585.1; JOINED.  
 DR EMBL: Y16524; CAA76272.1; -  
 DR EMBL: Y16525; CAA76272.1; JOINED.  
 DR EMBL: Y16526; CAA76272.1; JOINED.  
 DR EMBL: Y16527; CAA76272.1; JOINED.  
 DR EMBL: Y16528; CAA76272.1; JOINED.  
 DR EMBL: Y16529; CAA76272.1; JOINED.  
 DR EMBL: Y16530; CAA76272.1; JOINED.  
 DR EMBL: Y16531; CAA76272.1; JOINED.  
 DR EMBL: Y16532; CAA76272.1; JOINED.  
 DR MGD: MGI:23029; PIGR.  
 DR PFAM: PF00047; ig; 5.  
 KW Immunoglobulin domain; Repeat; Transmembrane; Glycoprotein; Signal.  
 FT SIGNAL 1 18 POTENTIAL.  
 FT CHAIN 19 771 POLYMERIC-IMMUNOGLOBULIN RECEPTOR.  
 FT CHAIN 19 611 SECRETORY COMPONENT.  
 FT DOMAIN 19 645 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 646 668 POTENTIAL.  
 FT DOMAIN 669 771 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 33 117 IG-LIKE V-TYPE DOMAIN 1.  
 FT DOMAIN 145 227 IG-LIKE V-TYPE DOMAIN 2.  
 FT DOMAIN 250 331 IG-LIKE V-TYPE DOMAIN 3.  
 FT DOMAIN 363 447 IG-LIKE V-TYPE DOMAIN 4.  
 FT DOMAIN 477 553 IG-LIKE V-TYPE DOMAIN 5.  
 FT DISULFID 40 110 POTENTIAL.  
 FT DISULFID 152 220 POTENTIAL.  
 FT DISULFID 257 324 POTENTIAL.  
 FT DISULFID 370 440 POTENTIAL.  
 FT DISULFID 484 546 POTENTIAL.  
 FT CARBOHYD 90 90 POTENTIAL.  
 FT CARBOHYD 147 147 POTENTIAL.  
 FT CARBOHYD 170 170 POTENTIAL.  
 FT CARBOHYD 206 206 POTENTIAL.  
 FT CARBOHYD 420 420 POTENTIAL.  
 FT CARBOHYD 471 471 POTENTIAL.  
 FT CONFLICT 159 159 A -> V (IN REF. 1).  
 FT CONFLICT 396 396 V -> A (IN REF. 1).  
 FT CONFLICT 620 620 G -> R (IN REF. 1).  
 SQ SEQUENCE 771 AA: 84998 MW: 78681302EC710730 CRC64;

Query Match 9.0%; Score 254; DB 1; Length 771;  
 Best Local Similarity 36.8%; Pred. No. 1.08e-25;  
 Matches 43; Conservative 26; Mismatches 42; Indels 6; Gaps 5;

Db 13 FSGVSTKSPFGQEVSSIEGDSVSTCYPTDTSVNRHTRKYWCRO-GASGMCCTLLSSN 71  
 QY 10 FLFVSGALRILPVRVKGELGSGVTKC--P-LP-EMHVIYLCRMAGSGTCTGVST 65  
 Db 72 GYLKSEYSGRANLNPENFTVNIETQDDTGSYKCGLG-TSNRGLSTFDSVLE 127  
 QY 66 NFIAEKYKGRVTLKQYPRKNLFVLTQTESDSGVYACGAGMTDRGKTQKTVLNV 122

RESULT 2  
 ID PIGR\_HUMAN STANDARD; PRT; 764 AA.

AC P01833;  
 DT 21-JUL-1986 (Rel. 01, Created)

DT 01-OCT-1986 (Rel. 34, Last sequence update)

DT 15-FEB-2000 (Rel. 39, Last annotation update)

DE DE POLYMERIC-IMMUNOGLOBULIN RECEPTOR PRECURSOR (POLY-IG RECEPTOR) (PIGR)  
 GN [CONTAINS: SECRETORY COMPONENT].  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92039621.  
 RA Krajcic P., Grzeschik K.H., Geurts van Kessel A.H., Olaisen B.,  
 RA Brandtzaeg P.;  
 RT "The human transmembrane secretory component (poly-Ig receptor):  
 RT molecular cloning, restriction fragment length polymorphism and  
 RT chromosomal sublocalization.";  
 RL Hum. Genet. 87:642-648(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92387236.  
 RA Krajcic P., Kvale D., Tasken K., Brandtzaeg P.;  
 RT "Molecular cloning and exon-intron mapping of the gene encoding human  
 RT transmembrane secretory component (the poly-Ig receptor).";  
 RL Eur. J. Immunol. 22:2309-2315(1992).  
 RN [3]  
 RP SEQUENCE OF 72-764 FROM N.A.  
 RX MEDLINE: 89149795.  
 RA Krajcic P., Solberg R., Sandberg M., Oyen O., Jahnsen T.,  
 RA Brandtzaeg P.;  
 RT "Molecular cloning of the human transmembrane secretory component  
 RT (poly-Ig receptor) and its mRNA expression in human tissues.";  
 RL Biochem. Biophys. Res. Commun. 156:783-789(1989).  
 RN [4]  
 RP SEQUENCE OF 19-577, DISULFIDE BONDS, AND CARBOHYDRATE-BINDING SITES.  
 RX MEDLINE: 85128981.  
 RA Eifert H., Quentin E., Decker J., Hillemair S., Hufschmidt M.,  
 RA Klingmüller D., Weber M.H., Hilschmann N.;  
 RT "The primary structure of human free secretory component and the  
 RT arrangement of disulfide bonds.";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 365:1489-1495(1984).  
 RN [5]  
 RP SEQUENCE OF 19-577.  
 RX MEDLINE: 91315750.  
 RA Eifert H., Quentin E., Wiederhold M., Hillemair S., Decker J.,  
 RA Weber M., Hilschmann N.;  
 RT "Determination of the molecular structure of the human free secretory  
 RT component.";  
 RL Biol. Chem. Hoppe-Seyler 372:119-128(1991).  
 RN [6]  
 RP SEQUENCE OF 118-138; 212-230; 232-268; 273-288 AND 578-603.  
 RX MEDLINE: 97379357.  
 RA Hughes G.J., Fruiter S., Savoy L.-A., Reason A.J., Morris H.R.,  
 RA Jaton J.-C.;  
 RT "Human free secretory component is composed of the first 595 amino  
 RT acid residues of the polymeric immunoglobulin receptor.";  
 RL FEBS Lett. 410:443-446(1997).  
 CC -!- FUNCTION: THIS RECEPTOR BINDS POLYMERIC IGA AND IGM AT THE  
 CC BASOLATERAL SURFACE OF EPITHELIAL CELLS. THE COMPLEX IS THEN  
 CC TRANSPORTED ACROSS THE CELL TO BE SECRETED AT THE APICAL SURFACE.  
 CC DURING THIS PROCESS A CLEAVAGE OCCURS THAT SEPARATE THE  
 CC EXTRACELLULAR (KNOWN AS THE SECRETORY COMPONENT) FROM THE  
 CC TRANSMEMBRANE SEGMENT.  
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ALSO SECRETED.  
 CC -!- SIMILARITY: CONTAINS 5 IMMUNOGLOBULIN-LIKE V-TYPE DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: S62403; AAB20203.1; -  
 DR EMBL: S43449; AAB23176.1; -

Best local similarity 55.0%;  
 41; Conservative  
 Matches 30; Mismatches 40; Indels 6; Gaps 4;  
 Seq. no. 2,348 25;



use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: X00412; CAA25118.1; -.  
PIR: A02111; QRRBG.  
PIR: A28077; A28077.  
PFAM: PF00047; ig; 5.  
Immunoglobulin domain; Repeat; Transmembrane; Glycoprotein; Signal; Polymorphism.  
FT SIGNAL 1 18  
FT CHAIN 19 773 POLYMERIC-IMMUNOGLOBULIN RECEPTOR.  
FT CHAIN 19 615 SECRETORY COMPONENT.  
FT DOMAIN 19 647 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 648 670 POTENTIAL.  
FT DOMAIN 671 773 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 30 136 IG-LIKE V-TYPE DOMAIN 1.  
FT DOMAIN 231 243 IG-LIKE V-TYPE DOMAIN 2.  
FT DOMAIN 244 350 IG-LIKE V-TYPE DOMAIN 3.  
FT DOMAIN 351 456 IG-LIKE V-TYPE DOMAIN 4.  
FT DOMAIN 457 558 IG-LIKE V-TYPE DOMAIN 5.  
FT DISULFID 46 115 POTENTIAL.  
FT DISULFID 155 225 POTENTIAL.  
FT DISULFID 260 324 POTENTIAL.  
FT DISULFID 369 438 POTENTIAL.  
FT DISULFID 478 538 POTENTIAL.  
FT CARBOHYD 108 108 IN ALLOTYPIC T61.  
FT CARBOHYD 418 418 IN ALLOTYPIC T62 (PARTIAL) AND T63.  
FT VARIANT 88 88 K -> N (IN ALLOTYPIC T61).  
FT VARIANT 94 94 D -> E (IN ALLOTYPIC T61).  
FT VARIANT 101 108 TVDLTON -> YLNLSOS (IN ALLOTYPIC T61).  
FT VARIANT 110 110 S -> T (IN ALLOTYPIC T63).  
SQ SEQUENCE 773 AA; 83886 MW; DF2G44D2F1193G65 CRC64;

Query Match 7.9%; Score 223; DB 1; Length 773;

Best Local Similarity 45.5%; Pred. No. 5.43e-20; Mismatches 11; Mismatches 30; Indels 7; Gaps 5;

Db 39 GDSVITCYPTTSVTRHSRKFRCRE-EESGRCVTLAST-GYTSQESYGRGKLTDFPDG 96  
Qy 30 GGSVTKC-PLPEM-HVRIYLCREMAGSGTCGTVTSTNFINKAEYKGRVTLKQYPRKN 85  
Db 97 EFVTVTLQNDQSGSKGCVGVN-GRG 123  
Qy 86 LFLVEVTQLTESDSGVYACGAGMNTDRG 113

RESULT 6 STANDARD; PRT; 102 AA.

ID PIGR\_PIG  
AC Q29244;  
DT 15-FEB-2000 (Rel. 39, Created)  
DT 15-FEB-2000 (Rel. 39, Last sequence update)  
DT 15-FEB-2000 (Rel. 39, Last annotation update)  
DE POLYMERIC-IMMUNOGLOBULIN RECEPTOR (POLY-IG RECEPTOR) (PIGR)  
DE (FRAGMENT).  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=SMALL INTESTINE;  
RX MEDLINE; 96327607.  
RA Winteroe A.K., Fredholm M., Davies W.;  
RT "Evaluation and characterization of a porcine small intestine cDNA library: analysis of 839 clones.";  
RL Mamm. Genome 7:509-517(1996).  
CC -!- FUNCTION: THIS RECEPTOR BINDS POLYMERIC IGA AND IGM AT THE BASOLATERAL SURFACE OF EPITHELIAL CELLS. THE COMPLEX IS THEN TRANSPORTED ACROSS THE CELL TO BE SECRETED AT THE APICAL SURFACE.  
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ALSO SECRETED.

-!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: F14851; CAA23294.1; -.  
Immunoglobulin domain; Repeat; Transmembrane; Glycoprotein.

FT NON\_TER 1 102  
FT NON\_TER 102 102  
SQ SEQUENCE 102 AA; 11205 MW; 82C915264B1508E8 CRC64;

Query Match 6.6%; Score 188; DB 1; Length 102;  
Best Local Similarity 38.3%; Pred. No. 8.21e-14;  
Matches 36; Conservative 19; Mismatches 31; Indels 8; Gaps 8;

Db 2 LOVLKPELYLDGSGVTFDCALQGMANVAFELCOLKNGK-TCNVVINTLG-KKAQD 59  
Qy 17 LRIL-PEVK-VEGELGSGVTIKCPL-PEM-HVRIYLCREMAGSGTCGTVTSTNFIRA-E 71

Db 60 PEGRILLTP-KENSHFSVHTGLRKEDAGHYLCG 92

Qy 72 YKGRVTLKQYPRKNLFLVEVTQLTESDSGVYACG 105

RESULT 7

ID CM35\_HUMAN STANDARD; PRT; 224 AA.  
AC Q08708;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE CMRF35 ANTIGEN PRECURSOR.  
GN CMRF35.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 92249405.  
RA Jackson D.G., Hart D.N.J., Starling G., Bell J.I.;  
RT "Molecular cloning of a novel member of the immunoglobulin gene superfamily homologous to the polymorphic immunoglobulin receptor.";  
RL Eur. J. Immunol. 22:1157-1163(1992).  
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (POTENTIAL).  
CC -!- TISSUE SPECIFICITY: PRESENT ON THE SURFACE OF MONOCYTES, NEUTROPHILS, A PROPORTION OF PERIPHERAL BLOOD T AND B LYMPHOCYTES AND LYMPHOCYTIC CELL LINES.  
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: X66171; CAA46948.1; -.  
DR PFAM: PF00047; ig; 1.  
DR Antigen; Transmembrane; Glycoprotein; Signal.  
KW SIGNAL 1 20 POTENTIAL.  
FT CHAIN 21 224 CMRF35 ANTIGEN.  
FT DOMAIN 21 183 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 184 204 POTENTIAL.  
FT DOMAIN 205 224 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 36 117 IG-LIKE V-TYPE DOMAIN.  
FT DOMAIN 123 175 PRO-RICH.  
FT DISULFID 43 110 POTENTIAL.  
FT DISULFID 57 65





DT 15-FEB-2000 (Rel. 39, Last annotation update)  
 DE CELL SURFACE A33 ANTIGEN PRECURSOR.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=COLON CARCINOMA;  
 RX MEDLINE; 97165045.  
 RA Heath J.K., White S.J., Johnstone C.N., Catimel B., Simpson R.J.,  
 RA Moritz R.L., Tu G.F., Ji H., Whitehead R.H., Groenen L.C.,  
 RA Scott A.M., Ritter G., Cohen L., Welt S., Old L.J., Nice E.C.,  
 RA Burgess A.W.;  
 RT "The human A33 antigen is a transmembrane glycoprotein and a novel  
 RT member of the immunoglobulin superfamily.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 94:469-474(1997).  
 RN [2]  
 RP POST-TRANSLATIONAL MODIFICATIONS.  
 RX MEDLINE; 97396159.  
 RA Ritter G., Cohen L.S., Nice E.C., Catimel B., Burgess A.W.,  
 RA Moritz R.L., Ji H., Heath J.K., White S.J., Welt S., Old L.J.,  
 RA Simpson R.J.;  
 RT "Characterization of posttranslational modifications of human A33  
 RT antigen, a novel palmitoylated surface glycoprotein of human  
 RT gastrointestinal epithelium.";  
 RL Biochem. Biophys. Res. Commun. 236:682-686(1997).  
 CC -!- FUNCTION: MAY PLAY A ROLE IN CELL-CELL RECOGNITION AND SIGNALING.  
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN NORMAL GASTROINTESTINAL  
 CC EPITHELIUM AND IN 95% OF COLON CANCERS.  
 CC -!- PTM: N-GLYCOSYLATED, CONTAINS APPROXIMATELY 8 K OF N-LINKED  
 CC CARBOHYDRATE.  
 CC -!- PTM: PALMITOYLATED.  
 CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.  
 CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; U79725; AAC50957.1; -;  
 DR HSP; P06907; INEU.  
 DR MIM; 602117; -;  
 DR PFAM; PF00047; ig; 2.  
 KW Immunoglobulin domain; Lipoprotein; Palmitate; Glycoprotein;  
 KW Transmembrane; Signal; Antigen.  
 FT SIGNAL 1 21  
 FT CHAIN 22 319 CELL SURFACE A33 ANTIGEN.  
 FT DOMAIN 22 235 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 236 256 POTENTIAL.  
 FT DOMAIN 257 319 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 22 138 IG-LIKE V-TYPE DOMAIN.  
 FT DOMAIN 139 229 IG-LIKE C2-TYPE DOMAIN.  
 FT DOMAIN 258 261 POLY-CYS.  
 FT DISULFID 146 222 POTENTIAL.  
 FT CARBOHYD 112 112 POTENTIAL.  
 FT CARBOHYD 200 200 POTENTIAL.  
 FT CARBOHYD 223 223 POTENTIAL.  
 SQ SEQUENCE 319 AA; 35632 MW; 9BFC7AAFA45C2408E CRC64; \*  
 Query Match 4.2%; Score 118; DB 1; Length 319;  
 Best Local Similarity 33.3%; Pred. No. 8.74e-03;  
 Matches 21; Conservative 16; Mismatches 23; Indels 3; Gaps 3;  
 Db 75 SNKNVINGELYKNRYSINNAEQSDASITIDQLTWADNGTYECSVLSMDELGWTKSRVR 134  
 QY 63 STTFIKAE-YKGRVTLKQYPRKNFLVEVTQLTESDSGVYACGAGMNTD-RGKTQ-KVT 119  
 Db 135 LLV 137

QY 120 LNV 122  
 RESULT 11  
 ID STP2\_RAT STANDARD; PRT; 116 AA.  
 AC P11101;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE NUCLEAR TRANSITION PROTEIN 2 (TP-2).  
 GN TNP2.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 89263797.  
 RA Luerksen H., Maier W.M., Hoyer-Fender S., Engel W.;  
 RT "The nucleotide sequence of rat transition protein 2 (TP2) cDNA.";  
 RL Nucleic Acids Res. 17:3585-3585(1989).  
 RN [2]  
 RP SEQUENCE OF 91-116.  
 RX MEDLINE; 87326384.  
 RA Cole K.D., Kistler W.S.;  
 RT "Nuclear transition protein 2 (TP2) of mammalian spermatids has a  
 RT very basic, carboxyl terminal domain.";  
 RL Biochem. Biophys. Res. Commun. 147:437-442(1987).  
 RN [3]  
 RP ZINC-BINDING.  
 RX MEDLINE; 92028893.  
 RA Baskaran R., Rao M.R.S.;  
 RT "Mammalian spermatid specific protein, TP2, is a zinc metalloprotein  
 RT with two finger motifs.";  
 RL Biochem. Biophys. Res. Commun. 179:1491-1499(1991).  
 CC -!- FUNCTION: IN THE ELONGATING SPERMATIDS OF MAMMALS, THE CONVERSION  
 CC OF NUCLEOSOMAL CHROMATIN TO THE COMPACT, NONNUCLEOSOMAL FORM FOUND  
 CC IN THE SPERM NUCLEUS IS ASSOCIATED WITH THE APPEARANCE OF A SMALL  
 CC SET OF BASIC CHROMOSOMAL TRANSITION PROTEINS.  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -!- SIMILARITY: STRONG, TO OTHER MAMMALIAN SPERMATID NUCLEAR  
 CC TRANSITION PROTEINS 2.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; X14776; CAA32882.1; -;  
 DR PIR; A26834; A26834.  
 DR PIR; S04094; S04094.  
 DR PFAM; PF01254; TP2; 1.  
 DR PROSITE; PS00970; TP2\_1; 1.  
 DR PROSITE; PS00971; TP2\_2; 1.  
 KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;  
 KW Nuclear protein; Zinc; Metal-binding.  
 FT DOMAIN 14 80 CYS/HIS-RICH.  
 SQ SEQUENCE 116 AA; 12994 MW; F80E505E1D8B5E9 CRC64;  
 Query Match 4.1%; Score 115; DB 1; Length 116;  
 Best Local Similarity 34.2%; Pred. No. 2.25e-02;  
 Matches 26; Conservative 21; Mismatches 24; Indels 5; Gaps 5;  
 Db 7 RGRAPTHPHSSSRPSQSHTNQACSHHCRSCGAGHPSSSSSPSPGPTKHKPTPMH 66  
 QY 292 RMALSSORPGSPRRSRQNNIYSCPRRAGDAAG-TGEAPVPGAPL-PP-APLQ 348  
 Db 67 SRYSPS-R-PSHRGSC 80  
 QY 349 VSESPWLHAPSLKTSK 364



```

Matches 17:  Conservative 15:  Mismatches 13:  Indels 1:  Gaps 1:
Db 75  GSNY-KTSLKSRVTSLSDFSKNLFSLKSSVTAADTAVYVCAGL 119
      :|||: 1: :|||: |||: : : :|||: 1: 1:
QY 63  STTFIKAEYKGRVTLKQYPRKNLFLEVLTQLTSDSGVYACGAGM 108

Search completed: Thu Aug 24 10:22:41 2000
Job time : 17 secs.

```



\*\*\*\*\*  
[W][A][T][E][R][S][E][R][E][S][E][R][E]  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Thu Aug 24 10:22:59 2000; MasPar time 32.75 Seconds  
Tabular output not generated. 825.707 Million cell updates/sec

Title: >US-09-135-238B-2  
Description: (1-390) from US09135238B.pep  
Perfect Score: 2830  
Sequence: 1 MDRWLWPLFLPVSGALRIL.....HOPAAAMEDSDSDDYINVPA 390

Scoring table: PAM 150  
Gap 11

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: sptrmbll2  
1:sp\_archaea 2:sp\_bacteria 3:sp\_fungi 4:sp\_human  
5:sp\_invertebrate 6:sp\_mammal 7:sp\_mhc 8:sp\_organelle  
9:sp\_phase 10:sp\_plant 11:sp\_rodent 12:sp\_unclassified  
13:sp\_vertebrate 14:sp\_virus

Statistics: Mean 49.245; Variance 97.409; scale 0.506

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Description	Pred. No.
1	2815	99.5	ANTI-FAS-INDUCED APOPT	0.00e+00
2	149	5.3	CMRF-35-H9.	1.01e-06
3	129	4.6	NUCLEAR TRANSITION PRO	8.53e-04
4	129	4.6	NUCLEAR TRANSITION PRO	8.53e-04
5	129	4.6	L2385.5 PROTEIN.	8.53e-04
6	126	4.5	KIAA0402 (FRAGMENT).	2.24e-03
7	126	4.5	KENDRIN.	2.24e-03
8	122	4.3	HYPOTHETICAL 23.8 KD P	7.99e-03
9	119	4.2	BCDNA.GH10711.	2.04e-02
10	115	4.1	KIAA0592 PROTEIN (FRAG	6.95e-02
11	112	4.0	EUS1 PROTEIN.	1.71e-01
12	113	4.0	LEUKOTRIENE B4 RECEPTO	1.27e-01
13	113	4.0	KOZAL1.1 PROTEIN.	1.27e-01
14	113	4.0	HYPOTHETICAL 53.3 KD P	1.27e-01
15	114	4.0	HYPOTHETICAL 91.0 KD P	9.40e-02
16	109	3.9	F08H9.2 PROTEIN.	4.15e-01
17	110	3.9	AMBIENTAL PH REGULATOR	3.10e-01
18	111	3.9	SIALIDASE (EC 3.2.1.18	2.31e-01
19	111	3.9	KIAA0973 PROTEIN (FRAG	2.31e-01
20	109	3.9	FATTY ACID SYNTHASE.	4.15e-01

21	108	3.8	289 10	Q9XF31	HYPOTHETICAL 31.7 KD P	5.56e-01
22	108	3.8	395 1	Q9Y918	395AA LONG HYPOTHETICA	5.56e-01
23	107	3.8	654 5	Q17982	SIMILARITY TO ERBB-3 R	7.43e-01
24	108	3.8	699 10	O81765	EXTENSIN-LIKE PROTEIN	5.56e-01
25	107	3.8	944 4	Q14163	KIAA0150 PROTEIN (FRAG	7.43e-01
26	107	3.8	1353 3	O74162	ICHI.	7.43e-01
27	108	3.8	3247 14	O65553	UL36.	5.56e-01
28	105	3.7	57 14	O55723	PUTATIVE SMALL BASIC P	1.32e+00
29	104	3.7	178 2	O9XAB5	PUTATIVE INTEGRAL MEMB	1.75e+00
30	106	3.7	416 10	O9XE67	HYPOTHETICAL 46.5 KD P	9.91e-01
31	105	3.7	508 10	O49444	COPPER-BINDING PROTEIN	1.32e+00
32	106	3.7	539 2	O83771	HYPOTHETICAL 63.1 KD P	9.91e-01
33	105	3.7	766 10	O39177	PEARL 4.	1.32e+00
34	104	3.7	776 11	Q92143	SEMAPHORIN W.	1.75e+00
35	104	3.7	815 5	O97214	L4830.10 PROTEIN.	1.75e+00
36	106	3.7	884 4	O00302	ERPROT 213-21.	9.91e-01
37	104	3.7	886 13	P70049	ORIGIN RECOGNITION COM	1.75e+00
38	106	3.7	897 11	O70495	PLENTY-OF-PROLINES-101	9.91e-01
39	103	3.6	126 3	Q12444	ORF YOR309C.	2.32e+00
40	103	3.6	279 2	O86663	PUTATIVE INTEGRAL MEMB	2.32e+00
41	102	3.6	662 5	Q21139	SIMILAR TO DROSOPHILA	3.07e+00
42	102	3.6	1179 4	O14120	DBP-5 NUCLEAR PROTEIN.	3.07e+00
43	102	3.6	1280 4	O75158	KIAA0670 PROTEIN (FRAG	3.07e+00
44	103	3.6	1336 14	O73550	POLYPROTEIN.	2.32e+00
45	103	3.6	1792 6	O46385	SUPERVILLIN.	2.32e+00

ALIGNMENTS

RESULT	ID	PRELIMINARY;	PRT;	390 AA.
1	O60667			
AC	O60667			
DT	01-AUG-1998 (Tremblrel. 07, Created)			
DT	01-AUG-1998 (Tremblrel. 07, Last sequence update)			
DT	01-MAY-1999 (Tremblrel. 10, Last annotation update)			
DE	ANTI-FAS-INDUCED APOPTOSIS.			
GN	TOSO.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
OC	Eutheria; Primates; Catarrhini; Hominidae; Homo.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 98246048.			
RA	HITOSHI Y., LORENS J., KITADA S.I., FISHER J., LABARGE M., RING H.Z.,			
RA	FRANCKE U., REED J.C., KINOSHITA S., NOLAN G.P.;			
RT	"Toso, a cell surface, specific regulator of Fas-induced apoptosis in			
RT	T cells.";			
RL	Immunity 8:461-471(1998).			
DR	EMBL; AF057557; AAC18830.1; -.			
DR	PFAM; PF00047; ig; 1.			
SQ	SEQUENCE 390 AA; 43146 MW; 15D1B5D8 CRC32;			

Query Match	99.5%;	Score 2815;	DB 4;	Length 390;
Best Local Similarity	99.7%;	Pred. No. 0.00e+00;		
Matches	389;	Conservative 0;	Mismatches 1;	Indels 0; Gaps 0;
Db	1	MDFWLWPLFLPVSGALRILPEVKVEGELGGVTKICPLPEMHVRIYLCREMGSGTGGT	60	
QY	1	MDRWLWPLFLPVSGALRILPEVKVEGELGGVTKICPLPEMHVRIYLCREMGSGTGGT	60	
Db	61	VVSTTNFIKAEYKGRVTLKQYPRKNLFLVEVTQLTSDSGVYACAGMNTDRGKTQVTL	120	
QY	61	VVSTTNFIKAEYKGRVTLKQYPRKNLFLVEVTQLTSDSGVYACAGMNTDRGKTQVTL	120	
Db	121	NHSEPEPWEQPMETPKFHLFQMPAYASSKSFVTRVTPPAQKGKVPVHHSSP	180	
QY	121	NHSEPEPWEQPMETPKFHLFQMPAYASSKSFVTRVTPPAQKGKVPVHHSSP	180	
Db	181	TQIITHRPVRVSSVAGDKPRTFLPSTTASISALEGLLKPTQPSYNNHTLHRQALD	240	
QY	181	TQIITHRPVRVSSVAGDKPRTFLPSTTASISALEGLLKPTQPSYNNHTLHRQALD	240	
Db	241	YGSQSGREGGFHILPTITLGLLALLGLVVKRAVERRKALSRARLRVRALESQ	300	

```

QY 241 YGSGGREGGPHILPILGLFLALLGLVYKRAVERKALSRARRLAVRWRALSSQ 300
|||||
Db 301 RPRGSPRPNQNIYSACPRRAGDAAGTGEAPVPGCAPLPAPLQVSESPWLHAPSL 360
|||||
QY 301 RPRGSPRPNQNIYSACPRRAGDAAGTGEAPVPGCAPLPAPLQVSESPWLHAPSL 360
|||||
Db 361 KTSCEYVSLYHQPAAAMEDSDSDYINVA 390
|||||
QY 361 KTSCEYVSLYHQPAAAMEDSDSDYINVA 390
|||||

RESULT 2
ID O95100 PRELIMINARY; PRT; 301 AA.
AC O95100;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
DE CMRF-35-H9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA GREEN B.J., CLARK G.J., HART D.N.J.;
RT "The CMRF-35 monoclonal antibody recognised a second leucocyte
RT membrane molecule with a domain similar to the poly Ig receptor.";
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF020314; AAD01646.1; -.
SQ SEQUENCE 301 AA; 33658 MW; E1BF9D11 CRC32;

Query Match 5.3%; Score 149; DB 4; Length 301;
Best Local Similarity 33.0%; Pred. No. 1.01e-06;
Matches 35; Conservative 21; Mismatches 42; Indels 8; Gaps 7;

Db 2 MLPWALLLVPCGF-ALSKCRTVAGPWG-SLSVOCPYKEHRTLNKYCRP-PQIFLCD 58
|||||
QY 4 WL-WPLVFLPVSGALRILPEVK-VEGELGSGVTIKPLPMH--VRILCREMAGSGTCG 59
|||||
Db 59 KIVETKSA-GKNGRVSIRDSPANLSFTVTLENTEEDAGTWCG 103
|||||
QY 60 TWSTTTFIKAEYKGRVTLKQYPRKNLFLIVEVTQLTESDSGVACG 105
|||||

RESULT 3
ID O64561 PRELIMINARY; PRT; 114 AA.
AC O64561;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1996 (TREMBlrel. 12, Last annotation update)
DE NUCLEAR TRANSITION PROTEIN 2 (TP-2).
GN TP2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-WISTAR; TISSUE=TESTES;
RA MEETEI A.R., RAO M.R.S.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: IN THE ELONGATING SPERMATIDS OF MAMMALS, THE CONVERSION
CC OF NUCLEOSOMAL CHROMATIN TO THE COMPACT, NONNUCLEOSOMAL FORM FOUND
CC IN THE SPERM NUCLEUS IS ASSOCIATED WITH THE APPEARANCE OF A SMALL
CC SET OF BASIC CHROMOSOMAL TRANSITION PROTEINS.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- SIMILARITY: STRONG, TO OTHER MAMMALIAN SPERMATID NUCLEAR
CC TRANSITION PROTEINS 2.
DR EMBL: U52958; AAB02693.1; -.
DR PROSITE: PS00970; TP2_1; 1.
DR PROSITE: PS00971; TP2_2; 1.
DR PFM: PF01254; TP2; 1.
KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
RN [1]

```

```

SQ SEQUENCE 114 AA; 12848 MW; 5BB5C17A CRC32;

Query Match 4.6%; Score 129; DB 11; Length 114;
Best Local Similarity 31.3%; Pred. No. 8.53e-04;
Matches 25; Conservative 27; Mismatches 22; Indels 6; Gaps 6;

Db 1 MDTKQSLPTTHPHSSSRPQSHNNOCASHCHSCSQAGHPSSSSSPGPTKHPK 60
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 289 LAVMRALSSQ-RPRGSPRPNQNIYSACPRRAGDAAG-TGEAPVPGCAPL-PP- 344
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Db 61 TPMHSRYSPS-R-PSHRGSC 78
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 345 APLQVSESPWLHAPSLKTSK 364
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 4
ID Q64390 PRELIMINARY; PRT; 115 AA.
AC Q64390;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1996 (TREMBlrel. 12, Last annotation update)
DE NUCLEAR TRANSITION PROTEIN 2 (TP-2).
GN TP2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE: 96341725.
RA SCHLUETER G., CELIK A., OBATA R., SCHLICKE M., HOFFERBERT S.,
RA "Schlung A., ADHAM I.M., ENGEL W.;
RT "Sequence analysis of the conserved protamine gene cluster shows that
RT it contains a fourth expressed gene.";
RL Mol. Reprod. Dev. 43:1-6(1996).
CC -!- FUNCTION: IN THE ELONGATING SPERMATIDS OF MAMMALS, THE CONVERSION
CC OF NUCLEOSOMAL CHROMATIN TO THE COMPACT, NONNUCLEOSOMAL FORM FOUND
CC IN THE SPERM NUCLEUS IS ASSOCIATED WITH THE APPEARANCE OF A SMALL
CC SET OF BASIC CHROMOSOMAL TRANSITION PROTEINS.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- SIMILARITY: STRONG, TO OTHER MAMMALIAN SPERMATID NUCLEAR
CC TRANSITION PROTEINS 2.
DR EMBL: Z46939; CAA87064.1; -.
DR PROSITE: PS00970; TP2_1; 1.
DR PROSITE: PS00971; TP2_2; 1.
DR PFM: PF01254; TP2; 1.
KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW Nuclear protein; Zinc; Metal-binding.
SQ SEQUENCE 115 AA; 12948 MW; 45AEE305 CRC32;

Query Match 4.6%; Score 129; DB 11; Length 115;
Best Local Similarity 31.3%; Pred. No. 8.53e-04;
Matches 25; Conservative 27; Mismatches 22; Indels 6; Gaps 6;

Db 1 MDTKQSLPTTHPHSSSRPQSHNNOCASHCHSCSQAGHPSSSSSPGPTKHPK 60
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 289 LAVMRALSSQ-RPRGSPRPNQNIYSACPRRAGDAAG-TGEAPVPGCAPL-PP- 344
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Db 61 TPMHSRYSPS-R-PSHRGSC 78
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 345 APLQVSESPWLHAPSLKTSK 364
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 5
ID O97197 PRELIMINARY; PRT; 876 AA.
AC O97197;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
DE L2385.5 PROTEIN.
GN L2385.5.
OS Leishmania major.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
RN [1]

```

[illegible]





```

Db 165 RTLICAPNYPNKHVPHLFEAITG-FLLPFLVAVVASYDIGRRQLQARRRRRR-TGR 222
| : | : : : | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 237 RALDYGSG-SRGEGGFHILPTILGILFLLALLGLVVK-RAVERR-KALS-RRARLAVR 292
| : | : : : | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 223 LVVL 226
: |
QY 293 MRAL 296
: |

RESULT 13
ID P90902 PRELIMINARY; PRT; 476 AA.
AC Q3Y4N4;
DT 01-MAY-1997 (TEMBLrel. 03, Created)
DT 01-MAY-1997 (TEMBLrel. 03, Last sequence update)
DT 01-NOV-1999 (TEMBLrel. 12, Last annotation update)
DE K02A11.1 PROTEIN.
DE K02A11.1 GN
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RA SWYE R.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAKINS T., HILLIER L., JER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
DR EMBL; Z75544; CAA99881.1; -.
DR HSSP; P55273; lbd8.
DR PFAM; PF00233; ank; 4.
SQ SEQUENCE 476 AA; 53687 MW; 2F755235 CRC32;

Query Match 4.0%; Score 113; DB 5; Length 476;
Best Local Similarity 35.2%; Pred. No. 1.27e-01;
Matches 19; Conservative 13; Mismatches 20; Indels 2; Gaps

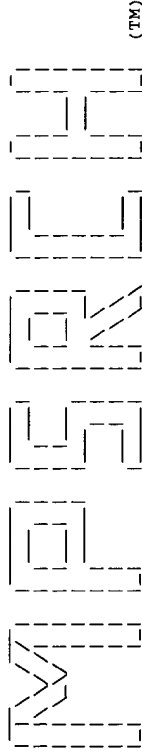
Db 6 LIAELPIERMTAQERTALARDRRRLQLR-RN-DERERALPPPRQRRLHFAA 57
| : | : : : | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 264 LLALLGLVVKRAVERRKALSRRARRLAVMRRALESSORPGSPRSPRNNIYSA 317
| : | : : : | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 14
ID Q3Y4N4 PRELIMINARY; PRT; 491 AA.
AC Q3Y4N4;
DT 01-NOV-1999 (TEMBLrel. 12, Created)
DT 01-NOV-1999 (TEMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TEMBLrel. 12, Last annotation update)
DE HYPOTHETICAL 53.3 KD PROTEIN (FRAGMENT).
DE DKFZP4340171.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominoidea; Hominidae; Homo.
RN [1]
SEQUENCE FROM N.A.
RX TISSUE-TESTIS.
RA POUSTKA A., KLEIN M., MEWES H.W., CASSENHUBER J., WIEMANN S.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL080183; CAB45765.1; -.
KW Hypothetical protein.

```



\*\*\*\*\*



\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

MPSrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Aug 24 10:23:55 2000; MasPar time 21.42 Seconds  
Tubular output not generated. 859.002 Million cell updates/sec

Title: >US-09-135-238B-2  
Description: (1-390) from US09135238B.pep  
Perfect score: 2830  
Sequence: 1 MDRWLWLPVLPVSGALRL.....HQPAMMEDSDSDYINVPA 390

Scoring table: PAM 150  
Gap 11  
Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: pir64  
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 49.586; Variance 110.938; scale 0.447

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query %	ID	Description	Pred. No.
1	252	8.9	764	1	QRUGS secretory component p	6.59e-21
2	245	8.7	757	1	S48841 secretory component p	8.34e-20
3	246	8.7	769	1	QRRTGS secretory component p	5.81e-20
4	244	8.6	757	2	I45956 polymeric immunoglobulin	1.20e-19
5	223	7.9	773	1	ORRBG secretory component p	2.17e-16
6	140	4.9	224	2	CMRF-35 antigen - hum	1.75e-04
7	136	4.8	519	2	S14529 transition protein 2	5.73e-04
8	119	4.2	303	2	A44215 EUS1 protein - equine	7.29e-02
9	115	4.1	116	2	S04094 transition protein 2	2.16e-01
10	115	4.1	179	2	C36801 hypothetical protein	2.16e-01
11	117	4.1	567	2	C69611 ABC transporter requi	1.26e-01
12	115	4.1	1353	2	T00347 hypothetical protein	2.16e-01
13	113	4.0	136	2	PN0536 Ig heavy chain v regi	3.68e-01
14	114	4.0	146	1	G1HUH2 Ig heavy chain precu	2.82e-01
15	113	4.0	491	2	T12501 hypothetical protein	3.68e-01
16	113	4.0	863	1	S65954 replication licensing	3.68e-01
17	110	3.9	113	2	S57667 transition protein 2	8.11e-01
18	109	3.9	125	2	S37455 Ig mu chain - human	1.05e-00
19	109	3.9	502	2	S36494 E2 protein - human pa	1.05e-00
20	111	3.9	913	2	S20590 exo-alpha-sialidase (	6.24e-01
21	109	3.9	2796	2	JC4743 fatty-acid synthase (	1.05e-00
22	107	3.8	95	2	E47624 Ig heavy chain v-II r	1.76e+00
23	108	3.8	252	2	C43679 ORF3 protein - Autogr	1.36e+00

24	108	3.8	395	2	B72478 hypothetical protein	1.36e+00
25	108	3.8	699	2	T05225 extensin homolog F17I	1.36e+00
26	107	3.8	1333	2	T00249 ichl protein - inky c	1.76e+00
27	105	3.7	57	2	T03060 probable small basic	2.93e+00
28	104	3.7	114	2	T01262 immunoglobulin heavy	3.76e+00
29	104	3.7	129	2	S44114 Ig heavy chain v regi	3.76e+00
30	105	3.7	133	2	S34010 Ig heavy chain v regi	2.93e+00
31	105	3.7	134	2	S54906 Ig heavy chain v regi	2.93e+00
32	104	3.7	208	2	T16953 hypothetical protein	3.76e+00
33	104	3.7	256	2	A75013 hypothetical protein	3.76e+00
34	105	3.7	348	1	OZ2QBK circumsporozoite prot	2.93e+00
35	104	3.7	483	2	S36470 E2 protein - human pa	3.76e+00
36	105	3.7	508	2	T04605 hypothetical protein	2.93e+00
37	106	3.7	539	2	H71280 hypothetical protein	2.27e+00
38	105	3.7	575	2	S58647 vacuolar transport pr	2.93e+00
39	104	3.7	1404	2	E36788 hypothetical protein	3.76e+00
40	103	3.6	126	2	S58321 probable membrane pro	4.83e+00
41	102	3.6	354	2	A48931 transmembrane glycopr	6.19e+00
42	102	3.6	876	1	A57988 regulatory protein ar	6.19e+00
43	102	3.6	1203	2	S26650 DNA-binding protein 5	6.19e+00
44	102	3.6	1280	2	T00365 hypothetical protein	6.19e+00
45	103	3.6	1792	2	T08878 supervillin p205 - bo	4.83e+00

ALIGNMENTS

RESULT 1

ENTRY QRUGS #type complete  
TITLE secretory component precursor - human  
ALTERNATE\_NAMES poly-Ig receptor; polymeric immunoglobulin receptor  
CONTAINS free secretory component; transmembrane secretory component  
ORGANISM #formal\_name Homo sapiens #common\_name man  
DATE 28-Aug-1985 #sequence\_revision 23-Aug-1996 #text\_change 16-Jul-1999

ACCESSIONS A46537; A55284; I38115; A32263; S38978; S13453; A02112  
REFERENCE A46537  
#authors Krajci, P.; Kvale, D.; Tasken, K.; Brandtzaeg, P.  
#journal Eur. J. Immunol. (1992) 22:2309-2315  
#title Molecular cloning and exon-intron mapping of the gene encoding human transmembrane secretory component (the poly-Ig receptor).  
#cross-references MUID:92387236  
#accession A46537  
#status not compared with conceptual translation  
#molecule\_type DNA  
#residues 1-764 #label KRA  
#cross-references GB:S43449; NID:g255097; PIDN:AAB23176.1; PID:g255098  
#experimental\_source leukocytes  
#note sequence extracted from NCBI backbone (NCBIP:113253)

REFERENCE A55284  
#authors Krajci, P.; Grzeschik, K.H.; Geurts van Kessel, A.H.; Olaisen, B.; Brandtzaeg, P.  
#journal Hum. Genet. (1991) 87:642-648  
#title The human transmembrane secretory component (poly-Ig receptor): molecular cloning, restriction fragment length polymorphism and chromosomal sublocalization.  
#cross-references MUID:92039621  
#accession A55284  
#molecule\_type mRNA  
#residues 1-764 #label KR2  
#cross-references GB:S62403; NID:g238235; PIDN:AAB20203.1; PID:g238236  
#experimental\_source colonic adenocarcinoma cell line  
#note sequence extracted from NCBI backbone (NCBIN:62403, NCBIP:62408)

REFERENCE I38115  
#authors Piskurich, J.F.; France, J.A.; Tamer, C.M.; Willmer, C.A.; Kaetzel, C.S.; Kaetzel, D.M.  
#journal Mol. Immunol. (1993) 30:413-421  
#title Interferon-gamma induces polymeric immunoglobulin receptor mRNA in human intestinal epithelial cells by a protein synthesis dependent mechanism.  
#cross-references MUID:93205018  
#accession I38115

```

##molecule_type mRNA
##residues 1-764 ##label RES
##cross-references EMBL:X73079; NID:9456345; PIDN:CAA51532.1;
PID:9456346
##note submitted to the EMBL/GenBank/DBJ databases by J.F.
Piskurich, February 1994

REFERENCE
#authors Krajci, P.; Solberg, R.; Sandberg, M.; Oyen, O.; Jahnsen, T.;
Brandtzaeg, P.
#journal Biochem Biophys. Res. Commun. (1989) 158:783-789
#title Molecular cloning of the human transmembrane secretory in
component (poly-Ig receptor) and its mRNA expression in
human tissues.
#cross-references MIM:89149795
#accession A32263
##molecule_type mRNA
##residues 72-764 ##label KR3
##cross-references GB:M24559; NID:9514365; PIDN:AAA36102.1; PID:9514366
S38978
#authors Falgreen-Gebauer, E.; Gebauer, W.; Bastian, A.; Kratzin,
H.D.; Eifert, H.; Zimmermann, B.; Karas, M.; Hilschmann,
N.
#journal Biol. Chem. Hoppe-Seyler (1993) 374:1023-1028
#title The covalent linkage of secretory component to IgA. Structure
of sigA.
#cross-references MIM:94121784
#accession S38978
##molecule_type protein
##residues 478-488;517-526;543-545 ##label PAL
##note disulfide bonds for unbound and IgA-bound forms
REFERENCE
#authors Eifert, H.; Quentin, E.; Wiederhold, M.; Hillemeir, S.;
Decker, J.; Weber, M.; Hilschmann, N.
#journal Biol. Chem. Hoppe-Seyler (1991) 372:119-128
#title Determination of the molecular structure of the human free
secretory component.
#cross-references MIM:91315750
#accession S13453
##molecule_type protein
##residues 19-135,'Q',137-157,'D',159-207,'DE',210-228,230-233,'N',
235-240,'Q',242-261,'Q',263-279,'N',281-391,'D',
393-499,'D',501-577 ##label E12
REFERENCE
#authors A02112
Eifert, H.; Quentin, E.; Decker, J.; Hillemeir, S.;
Hufschmidt, M.; Klingmuller, D.; Weber, M.H.; Hilschmann,
N.
#journal Hoppe-Seyler's Z. Physiol. Chem. (1984) 365:1489-1495
#title The primary structure of the human free secretory component
and the arrangement of the disulfide bonds.
#cross-references MIM:85128981
#accession A02112
##molecule_type protein
##residues 19-157,'D',159-207,'DE',210-228,230-233,'N',235-240,'Q',
242-261,'Q',263-279,'N',281-391,'D',393-499,'D',
501-577 ##label E1F
##note paper in German with English abstract
COMMENT As a 100K transmembrane receptor for polymeric immunoglobulins,
secretory component is expressed on secretory epithelium
basolateral membranes. Upon binding IgA or IgM, it becomes
phosphorylated, forms interchain disulfide bonds, undergoes
proteolysis and transcytosis. Free secretory component may
protect polymeric immunoglobulins from proteolysis in mucosal
excretions.
GENETICS
#gene GDB:PTGR
##cross-references GDB:120290; OMIM:173880
#map_position 1q31-1q41
#introns 15/1; 130/1; 349/1; 460/1; 569/1; 629/2; 670/1; 714/1; 733/3
#note the first intron occurs before the initiator codon
monomeric as a transmembrane receptor or free in mucosal
secretions; heterodimer composed of one chain of
secretory component, one chain of immunoglobulin J chain,
and two IgA heterotetramers; hetero-22-mer composed of one
chain of secretory component, one chain of immunoglobulin J
chain, and five IgM heterotetramers
#superfamily secretory component; immunoglobulin homology
duplication; glycoprotein; immunoglobulin receptor;
phosphoprotein; transcytosis; transmembrane protein
#domain signal sequence #status predicted #label SIG\
#product transmembrane secretory component #status
#predicted #label MATM\
#product free secretory component #status experimental
#label MATM\
#domain immunoglobulin homology #label IM1\
#domain immunoglobulin homology #label IM2\
#domain immunoglobulin homology #label IM3\
#domain immunoglobulin homology #label IM4\
#domain immunoglobulin homology #label IM5\
#domain transmembrane #status predicted #label TM\
#domain intracellular #status predicted #label INT\
#disulfide_bonds #status experimental\
#binding_site carbohydrate (Asn) (covalent) #status
experimental\
#disulfide_bonds (in Ig-unbound form) #status
experimental\
#disulfide_bonds interchain (to IgA alpha-1 chain-192)
#status experimental\
#binding_site cysteine (Cys) (covalent) (in Ig-bound
form) #status experimental\
#cleavage_site Lys-Ala (unidentified proteinase) #status
experimental\
#binding_site phosphate (Ser) (covalent) #status
Predicted
SUMMARY #length 764 #molecular_weight 83313 #checksum 8299
Query Match 8.9%; Score 252; DB 1; Length 764;
Best Local Similarity 35.0%; Pred. No. 6.59e-21;
Matches 41; Conservative 30; Mismatches 40; Indels 6; Gaps 4;
Db 13 FPAISKSPFPGPEVNSVSTCYPTVNRHTRKYWCRO-GARGSGITLISSE 71
QY 10 FLVPYSGALILDEPKYGEGLGSGVTIKC---PLP-EMHVRILYLCREMGSGTCGVVSTT 65
Db 72 GYVSSKYAGRANLTFNPGNTFVNTAQLSQDSDSGRYKCGLGINS-RLGSFDSVLEY 127
QY 56 NFIKAEYKGRVTLKQYPRKNLFLVEVTQLTESDSGVYACGAGMNTDRGKTQKVTLVN 122
RESULT 2
ENTRY S48841 #type complete
TITLE secretory component precursor - bovine
ALTERNATE_NAMES poly-Ig receptor; polymeric immunoglobulin receptor
CONTAINS free secretory component; transmembrane secretory component
ORGANISM #formal_name Bos primigenius taurus #common_name cattle
DATE 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change
10-Sep-1999
ACCESSIONS S48841
REFERENCE S48841
#authors Vermeer, H.; Warner, G.W.; de Boer, H.A.; Verbeet, M.P.H.
#submission submitted to the EMBL Data Library, September 1994
#description The cloning, tissue specific expression and interspecies
sequence comparison of the bovine polymeric immunoglobulin
receptor.
#accession S48841
#status preliminary
#molecule_type mRNA
#residues 1-757 #label VER
#cross-references EMBL:X81371; NID:9563340; PIDN:CAA57136.1;
PID:9563341
CLASSIFICATION #superfamily secretory component; immunoglobulin homology

```

KEYWORDS duplication; glycoprotein; immunoglobulin receptor; phosphoprotein; transcytosis; transmembrane protein

FEATURE 1-18 #product signal sequence #status predicted #label SIG\ 19-572 #product transmembrane secretory component #status predicted #label MATM\ 19-572 #product free secretory component #status predicted #label MATM\ 33-112 #domain immunoglobulin homology #label IM1\ 145-222 #domain immunoglobulin homology #label IM2\ 250-328 #domain immunoglobulin homology #label IM3\ 365-444 #domain immunoglobulin homology #label IM4\ 476-547 #domain immunoglobulin homology #label IM5\ 631-653 #domain transmembrane #status predicted #label TM\ 654-757 #domain intracellular #status predicted #label INT\ 40-110,56-64, 152-220,166-173, 257-324,271-279, 370-440,384-394, 481-543,495-502 83,420,468

665 #disulfide\_bonds #status predicted\ #binding\_site carbohydrate (Asn) (covalent) #status predicted\ #binding\_site phosphate (Ser) (covalent) #status predicted\ #length 757 #molecular-weight 82450 #checksum 7998

SUMMARY Query Match 8.7%; Score 245; DB 1; Length 757; Best Local Similarity 36.7%; Pred. No. 8.34e-20; Matches 40; Conservative 25; Mismatches 37; Indels 7; Gaps 5;

Db 25 PE-EVSSVEGRSVKCYPPPTSVNRHTRKYWCRO-GAQCCTTLISSEGYSDDYVGRA 82  
 QY 21 PEVKVEGLGSGVTIKC--PLP-EMHVRILCREMAGSGTCGTVTTFIAEKYGRV 76  
 Db 83 NLTFNPESGTFVVDISHTHDKDSGRYKGLGISS-RGLNFDVSLEVSQD 130  
 QY 77 TLKQYPRKNLFLVEVTQLTSDSGYACGAGMNTDRGKTQKVTNLNVHSE 125

RESULT 3 QRTGS #type complete ENTRY secretory component precursor - rat TITLE poly-ig receptor; polymeric immunoglobulin receptor CONTAINS free secretory component; transmembrane secretory component ORGANISM #formal\_name Rattus norvegicus #common\_name Norway rat DATE 07-Sep-1990 #sequence\_revision 23-Aug-1996 #text\_change 16-Jul-1999

ACCESSIONS S05407; S54731 REFERENCE S05407 Banting, G.; Brake, B.; Braghetta, P.; Luzio, J.P.; Stanley, K.K. FEBS Lett. (1989) 254:177-183 Intracellular targeting signals of polymeric immunoglobulin receptors are highly conserved between species. #cross-references MUI:89378226 #accession S05407 #molecule\_type mRNA #residues 1-769 #label BAN #cross-references EMBL:X15741; NID:g56464; PIDN:CAA33758.1; PID:g56465 monomeric as a transmembrane receptor or free in mucosal secretions; heterodimer composed of one chain of secretory component, one chain of immunoglobulin J chain, and two IgA heterotetramers; hetero-22-mer composed of one chain of secretory component, one chain of immunoglobulin J chain, and five IgM heterotetramers

CLASSIFICATION #superfamily secretory component; immunoglobulin homology duplication; glycoprotein; immunoglobulin receptor; phosphoprotein; transcytosis; transmembrane protein

KEYWORDS #domain signal sequence #status predicted #label SIG\ 1-18 #product transmembrane secretory component #status predicted #label MATM\ 19-582 #product free secretory component #status predicted

33-112 #label MATF\ 145-222 #domain immunoglobulin homology #label IM1\ 250-328 #domain immunoglobulin homology #label IM2\ 363-442 #domain immunoglobulin homology #label IM3\ 477-548 #domain immunoglobulin homology #label IM4\ 644-666 #domain immunoglobulin homology #label IM5\ 667-769 #domain transmembrane #status predicted #label TM\ 40-110,56-64, 152-220,370-440, 384-394,484-546, 498-505 90,135,471

488-522 #disulfide\_bonds #status predicted\ 488 #binding\_site carbohydrate (Asn) (covalent) #status predicted\ #disulfide\_bonds interchain (to IgA alpha-1 chain-192) #status predicted\ 522 #binding\_site cysteine (Cys) (covalent) (in Ig-bound form) #status predicted\ 678 #binding\_site phosphate (Ser) (covalent) #status predicted\ #length 769 #molecular-weight 84798 #checksum 7578

SUMMARY Query Match 8.7%; Score 246; DB 1; Length 769; Best Local Similarity 35.0%; Pred. No. 5.81e-20; Matches 41; Conservative 30; Mismatches 40; Indels 6; Gaps 5;

Db 13 FSGYSTOSPFGPODVSSIEGNSVITCYPPDTSVNRHTRKYWCRO-GANGYCALIISN 71  
 QY 10 FLPVSGALRILPEKVEGLGSGVTIKC--P-LP-EMHVRILCREMAGSGTCGTVVST 65  
 Db 72 GYLSKEYSGRASLNFENSTFVINTAHLQEDTGSYKGLG-TTNRGLFFDVSLEV 127  
 QY 66 NFIAEKYGRVTLKQYPRKNLFLVEVTQLTSDSGYACGAGMNTDRGKTQKVTNLNV 122

RESULT 4 ENTRY #type complete TITLE polymeric immunoglobulin receptor - bovine ORGANISM #formal\_name Bos primigenius taurus #common\_name cattle DATE 19-Dec-1997 #sequence\_revision 19-Dec-1997 #text\_change 23-Jul-1999

ACCESSIONS I45956 REFERENCE I45956 Kulseth, M.A.; Krajci, P.; Myklebost, O.; Rogne, S. DNA Cell Biol. (1995) 14:251-256 Cloning and characterization of two forms of bovine polymeric immunoglobulin receptor cDNA. #cross-references MUI:95186063 #accession I45956 #status preliminary; translated from GB/EMBL/DBJ #molecule\_type mRNA #residues 1-757 #label KUL #cross-references GB:I04797; NID:g388279; PIDN:AAC41620.1; PID:g388280 CLASSIFICATION #superfamily secretory component; immunoglobulin homology FEATURE 145-222 #domain immunoglobulin homology #label IMM 145-222 #length 757 #molecular-weight 82434 #checksum 8761

SUMMARY Query Match 8.6%; Score 244; DB 2; Length 757; Best Local Similarity 37.0%; Pred. No. 1.20e-19; Matches 37; Conservative 24; Mismatches 33; Indels 6; Gaps 4;

Db 33 GBSVSIKCYKYPPTSVNRHTRKYWCRO-GAQCCTTLISSEGYSDDYVGRANLTFPESG 91  
 QY 30 GGSVTIKC---PLP-EMHVRILCREMAGSGTCGTVVSTTFNFIKAEYKGRVTLKQIPRN 85  
 Db 92 FTVVDISHLTHDKDSGRYKGLGISS-RGLNFDVSLEVSQD 130  
 QY 86 LPLVEVTQLTSDSGYACGAGMNTDRGKTQKVTNLNVHSE 125

RESULT 5

```

ENTRY
  QRRBC      #type complete
  TITLE      Secretory component precursor - rabbit
  ALTERNATE_NAMES
  CONTAINS   free secretory component; transmembrane secretory component
  ORGANISM   #formal_name Oryctolagus cuniculus #common_name domestic
              rabbit
  DATE       15-Nov-1984 #sequence_revision 15-Nov-1984 #text_change
              16-Jul-1999
  ACCESSIONS A02111; A28077
  REFERENCE  A02111
  #authors   Mostov, K.E.; Friedlander, M.; Blobel, G.
  #journal   Nature (1984) 308:37-43
  #title     The receptor for transendothelial transport of IgA and IgM
              contains multiple immunoglobulin-like domains.
  #cross-references M01D:8412246
  #accession A02111
  #molecule_type mRNA
  #residues   1-773 #label MOS
  #cross-references GB:X00412; GB:X01291; NID:g1595; PIDN:CAA25118.1;
              PID:g1596
  #note      the authors translated the codon ACC for residue 54 as
              Asn
  REFERENCE  A28077
  #authors   Frutiger, S.; Hughes, G.J.; Hanly, W.C.; Jaton, J.C.
  #journal   J. Biol. Chem. (1988) 263:8120-8125
  #title     Rabbit secretory components of different allotypes vary in
              their carbohydrate content and their sites of N-linked
              glycosylation.
  #cross-references M01D:88228032
  #accession A28077
  #molecule_type protein
  #residues   87-114:410-424 #label FRU
  COMMENT    This receptor binds polymeric IgA and IgM at the basolateral
              surface of epithelial cells. The complex is then transported
              across the cell to be secreted at the apical surface. During this
              process, cleavage occurs to separate the extracellular portion,
              also known as the secretory component, from the transmembrane
              segment.
  COMMENT    The five domains exhibit homology with immunoglobulin V regions.
              The similarity is strongest between the fourth domain and kappa
              chain V regions.
  COMMENT    Alternative splicing in the extracellular domain leads to high or
              low molecular weight forms of secretory component.
  CLASSIFICATION #superfamily secretory component; immunoglobulin homology
  KEYWORDS       alternative splicing; duplication; glycoprotein;
              immunoglobulin receptor; polymorphism; transcytosis;
              transmembrane protein
  FEATURE
    1-18        #domain signal sequence #status predicted #label SIG\
    19-773      #product transmembrane secretory component #status
              predicted #label MATM\
    19-575      #product free secretory component #status predicted
              #label MATF\
    30-647      #domain extracellular #status predicted #label EXT\
    39-117      #domain immunoglobulin homology #label IM1\
    148-227     #domain immunoglobulin homology #label IM2\
    253-326     #domain immunoglobulin homology #label IM3\
    362-440     #domain immunoglobulin homology #label IM4\
    471-540     #domain immunoglobulin homology #label IM5\
    648-670     #domain transmembrane #status predicted #label TMW\
    671-773     #domain intracellular #status predicted #label INT\
    46-115,155-225,
    260-324,359-438,
    478-538     #disulfide_bonds #status predicted\
    418         #binding_site carbohydrate (Asn) (covalent) (partial)
              #status experimental\
              #binding_site carbohydrate (Asn) (covalent) #status
              experimental
  SUMMARY      #length 773 #molecular-weight 83886 #checksum 7723
  Query Match 7.9%; Score 223; DB 1; Length 773;
  Best Local Similarity 45.5%; Pred. No. 2.17e-16;
  Matches 40; Conservative 11; Mismatches 30; Indels 7; Gaps 5;

```

```

Db 39 GDSVSIICYPPTSVTRHSRKFRCRE-ESGHCVTLAST-GYTSEYSGRGLTDFPKG 96
QY 30 GGSVTIKC--PIPEM--HVRIYICREMGSGCGTGVVSTTTFIKAEYGRVTLKQYPRKN 85
Db 97 EFVTVQQLTQNDSGSYKCGVGN-GRG 123
QY 86 LFLVEVTQLTESDSGYACGAGMNTDRG 113

RESULT      6
ENTRY       I37243      #type complete
TITLE       CRP-35 antigen - human
ORGANISM    #formal_name Homo sapiens #common_name man
DATE        12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change
              28-Feb-1997
ACCESSIONS  I37243
REFERENCE    I37243
#authors     Jackson, D.G.; Hart, D.N.; Starling, G.; Bell, J.I.
#journal     Eur. J. Immunol. (1992) 22:1157-1163
#title       Molecular cloning of a novel member of the immunoglobulin
              gene superfamily homologous to the polymeric immunoglobulin
              receptor.
#cross-references M01D:92249405
#accession   I37243
#status      preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues    1-224 #label RES
#cross-references EMBL:X66171; NID:g396169; PID:g396170
GENETICS
#gene        CMRF35
SUMMARY      #length 224 #molecular-weight 24830 #checksum 8067
              Query Match 4.9%; Score 140; DB 2; Length 224;
              Best Local Similarity 29.7%; Pred. No. 1.75e-04;
              Matches 30; Conservative 23; Mismatches 44; Indels 4; Gaps 3;
Db 13 ALLLLVPGVPLSHPMVAGVGSISVOCKEHEHTLNKEWCRP-PQILRCDKIVET 71
QY 7 PLFLPVSAGAILPEVKEGSGVTIKCLPEMH--VRIYICREMGSGTGVVST 64
Db 72 KGSAG-GRKGRVSRIDSPANLSFTVTLENTPEADAGTYWCG 111
QY 65 TNFIKAEYGRVTLKQYPRKNLFLVEVTQLTESDSGYACG 105

RESULT      7
ENTRY       S14529      #type complete
TITLE       transition protein 2 - mouse
ORGANISM    #formal_name Mus musculus #common_name house mouse
DATE        20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
              16-Jul-1999
ACCESSIONS  S14529; JQ0801; A28447
REFERENCE    S14529
#authors     Nelki, S.; Greenhall, C.; Shanahan, F.; Dudley, K.
#journal     submitted to the EMBL Data Library, November 1989
#description Identification and DNA sequence of a family of related cDNA's
              one of which is haploid specific and has complete homology.
#accession   S14529
#status      preliminary
#molecule_type mRNA
#residues    1-519 #label NEL
#cross-references EMBL:X17068; NID:g54873; PIDN:CAA34914.1; PID:g54874
REFERENCE    JQ0801
#authors     Kleene, K.C.; Gerstel, J.; Shih, D.
#journal     Gene (1990) 95:301-302
#title       Nucleotide sequence of the gene encoding mouse transition
              protein 2.
#cross-references M01D:91065546
#accession   JQ0801
#molecule_type DNA
#residues    403-512, 'T', 514-519 #label KLE
#cross-references GB:M60254; GB:M36193; NID:g202107; PIDN:AAA0469.1;

```

2

```

#authors      Telford, E.A.R.; Watson, M.S.; McBride, K.; Davison, A.J.
#submission   submitted to GenBank, March 1992
#description   The DNA sequence of equine herpesvirus-1.
#accession    E36801
#molecule_type DNA
##residues    1-179 ##label TEL
##cross-references GB:M85664; NID:g330791; PID:g330850
REFERENCE
A41831
#authors      Telford, E.A.R.; Watson, M.S.; McBride, K.; Davison, A.J.
#journal      Virology (1992) 189:304-316
#title        The DNA sequence of equine herpesvirus-1.
#cross-references M85664; NID:g330791; PID:g330850
#note         annotation; possible protein-coding frames
#contents     neither protein nor nucleotide sequence is given
GENETICS
#gene
#summary
SUMMARY
#length 179 #molecular-weight 19452 #checksum 4232

Query Match      4.1%; Score 115; DB 2; Length 179;
Best Local Similarity 43.5%; Pred. No. 2.16e-01;
Matches 20; Conservative 9; Mismatches 14; Indels 3; Gaps 3;

Db 88 LGALAM-AAQRKSGIRGPTVPVNAAMGSRQSRGRDGNPRGE 132
Qy 268 LGLVYKRAVERRRALSRARLAVRM-RALESSQRPRGSPR-PRSQ 311
||||| |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 11
ENTRY
TITLE      ABC transporter required for expression of cytochrome b4
            (ATP-) cydC - Bacillus subtilis
ORGANISM    05-Dec-1997 #sequence-revision 05-Dec-1997 #text_change
DATE        28-May-1999
C69611
A69580
#authors
Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
Daniel, R.A.; Denizot, F.; Devigne, K.M.; Duesterhoft, A.;
Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.;
Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
Guiseppi, G.; Guy, B.J.; Haga, K.; Haiech, J.; Harwood,
C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
Kashara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
Y.; Koetter, P.; Konlingstein, G.; Krogh, S.; Kumano, M.;
Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Mauel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
V.; Pohl, T.M.; Portetalle, D.; Porwollik, S.; Prescott, G.;
A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, E.;
Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
Takamaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
Terpstra, P.; Tognoni, F.; Tosato, V.; Uchiyama, S.;
Vandenbol, M.; Vannier, F.; Vassartotti, A.; Viari, A.;
Wambutt, R.; Wedler, E.; Wedler, H.; Weitznegger, T.;
Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumbstein, E.;
Toshikawa, H.; Danchin, A.
Nature (1997) 390:249-256
The complete genome sequence of the Gram-positive bacterium

```

```

#cross-references M85664; NID:g330791; PID:g330850
#accession    C69611
#status       preliminary; nucleic acid sequence not shown;
              translation not shown
#molecule_type DNA
##residues    1-567 ##label KUN
##cross-references GB:Z99123; GB:AL009126; NID:g2636240; PID:el186373;
              PID:g2636409
##experimental_source strain 168
GENETICS
#gene         cydC
CLASSIFICATION #superfamily ATP-binding cassette homology
KEYWORDS       ATP; p-loop
FEATURE        343-537
               360-367
SUMMARY        #domain ATP-binding cassette homology #label ABC\
               #region nucleotide-binding motif A (p-loop)
               #length 567 #molecular-weight 62805 #checksum 9210

Query Match      4.1%; Score 117; DB 2; Length 567;
Best Local Similarity 31.3%; Pred. No. 1.26e-01;
Matches 20; Conservative 21; Mismatches 20; Indels 3; Gaps 3;

Db 158 ILVRAAMPILIIIFMILGLVQAQRKADROWKSYQRLSNHFVDSLRGLT-LRFLGLSKSHK 216
Qy 254 ILPTILGLFL-LALLGLVWKAVER-RKALSRARLAVRMRALESSQRPRGSPRPSQ 311
||||| |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

Db 217 NIFY 220
Qy 312 NNII 315

RESULT 12
ENTRY
TITLE      T00347
              #type fragment
              hypothetical protein DKFzp566G1246.1, version 1 - human
              (fragment)
ALTERNATE_NAMES
ORGANISM     hypothetical protein KIAA0592
              #formal_name Homo sapiens #common_name man
DATE         01-Feb-1999 #sequence-revision 01-Feb-1999 #text_change
              13-Aug-1999
ACCESSIONS   T00347
REFERENCE     Z14086
#authors      Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani,
              H.; Nomura, N.; Ohara, O.
#journal      DNA Res. (1996) 5:31-39
#title        Prediction of the coding sequences of unidentified human
              genes. IX. The complete sequences of 100 new cDNA clones
              from brain which can code for large proteins in vitro.
#cross-references M85664; NID:g2636240; PID:el186373;
#accession    T00347
#status       translated from GB/EMBL/DBJ
#molecule_type mRNA
##residues    1-1353 ##label NAG
##cross-references EMBL:AB011164; NID:d1185434; PID:d1026448
##experimental_source brain; clone HJ2807
GENETICS
#note         KIAA0592
#summary      #length 1353 #checksum 6126

Query Match      4.1%; Score 115; DB 2; Length 1353;
Best Local Similarity 27.4%; Pred. No. 2.16e-01;
Matches 43; Conservative 40; Mismatches 66; Indels 8; Gaps 7;

Db 976 NPAALLP-TAASQISEVPKLPVLPAPSPSEHRRSHGLSVPLPGSGAGVSFD-L-PAQ 1032
Qy 200 KPTIFLPTTASQISGLGLKPKQTPSYNHHRLHQRLDYGQSGREGGQPHILPTI 259
||||| |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

Db 1033 ADLHSANKSRVYMRG--KRPQTRARARLAQESSEAEDEMSIPRPIQAQWDAISPNQ 1090
Qy 260 LGLFLLALLGLVWKAVERRRALSRARLAVRMRA-LESSQRPRGSPRPSQNNIYSAC 318
|||||
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

Db 1091 HRPQURASGEDSTEEALAAAFWEGGVPVGVDTSP 1127
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

```



```

Matches      17; Conservative      15; Mismatches      13; Indels      1; Gaps      17

Db    75 GSTNY-KTSLKSRVTISLDTSKNLFSLKLSVTAADTAVYVCARGL 119
      :|||: | : |||: |||| : : | : ||| | : | :
Qy    63 STTFIKAEYKGRVTLTKQYPRKNFLFVEVTLQTLESDSGYVACGAGM 108
      :|||: | : |||: |||| : : | : ||| | : | :

RESULT      15
ENTRY       TL2501          #type fragment
TITLE       hypothetical protein DKFPz4340171.1 - human (fragment)
ORGANISM    Homo sapiens   #common_name man
DATE        23-Jul-1999   #sequence_revision 23-Jul-1999 #text_change
                        23-Jul-1999
ACCESSIONS  TL2501
REFERENCE   Z17525
#authors    Poustka, A.; Klein, M.; Mewes, H.W.; Gassenhuber, J.;
            Wiemann, S.
#submission submitted to the Protein Sequence Database, June 1999
#acccession TL2501
#status     preliminary
#molecule_type mRNA
#residues   1-491 #label POU
#crosso-references EMBL:AL080183
#experimental_source adult testis; clone DKFPz4340171

GENETICS
#note       DKFPz4340171.1
SUMMARY     #length 491 #checksum 7607

Query Match      4.0%; Score 113; DB 2: Length 491;
Best Local Similarity 27.4%; Pred.No. 3.68e-01;
Matches 43; Conservative 40; Mismatches 66; Indels 8; Gaps 7

Db    114 NPAALLP-TAASQISEKVPLVPLELAFFPSSEHRRSHGLSEVPVLPGSGEAGVSFD-L-PAQ 170
      :|||: | : |||: |||| : : | : ||| | : | :
Qy    200 KPRFLPSTTASKISALEGLLKPKQTPSVNNHTLRHLRQRALDYGSGREGQGPHILPTI 259
      :|||: | : |||: |||| : : | : ||| | : | :

Db    171 ADTHLSANKSRVKRG--KRPPQTRARRLAAQSSTEDMSVPRGPIAWADCAISPNG 228
      :|||: | : |||: |||| : : | : ||| | : | :
Qy    260 LGLEFLLALLGVVRAVERRKALSRRARRAVRVRA-LESSQRGRGSPRPQRSQNIYSAC 318
      :|||: | : |||: |||| : : | : ||| | : | :

Db    229 HRPOLRAASGEDSTEELAAAAAAPWEGGPVPGVDPSR 265
      :|||: | : |||: |||| : : | : ||| | : | :
Qy    319 PRRA-RGADAAGTGAPVPGGAPLPAPIQ-VSESP 353
      :|||: | : |||: |||| : : | : ||| | : | :

Search completed: Thu Aug 24 10:24:19 2000
Job time : 24 secs.
```



\*\*\*\*\*  
WQSELEH  
\*\*\*\*\* (TW)

\*\*\*\*\*  
Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd  
MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Thu Aug 24 10:24:36 2000; Maspar time 12.26 Seconds  
Tabular output not generated.  
753.292 Million cell updates/sec

Title: >US-09-135-238B-2  
Description: (1-390) from US09135238B.pep  
Perfect Score: 2830  
Sequence: 1 MDRWLWPLFLVPLVSGALRL.....HQPAMMEDSDSDYINVPA 390  
Scoring table: PAM 150  
Gap 11  
Searched: 188963 segs, 23686106 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries  
Database: a-geneseq36  
1:geneseqp  
Statistics: Mean 36.230; Variance 187.638; scale 0.193

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	2815	99.5	390	1 Y05001	Human PIGRL-1 protein	5.91e-218
2	787	27.8	107	1 Y05002	Human PIGRL-1 protein	3.36e-51
3	254	9.0	771	1 W03180	Mouse poly-immunoglobulin	2.06e-09
4	252	8.9	607	1 W95601	Human secretory immunoglobulin	2.90e-09
5	247	8.7	746	1 W03178	Human poly-immunoglobulin	6.77e-09
6	245	8.7	757	1 W03179	Bovine poly-immunoglobulin	9.49e-09
7	246	8.7	769	1 W03181	Rat poly-immunoglobulin	8.01e-09
8	226	8.0	532	1 W50033	Human immunity related	2.32e-07
9	223	7.9	584	1 R14670	Truncated poly Ig-rece	3.83e-07
10	223	7.9	773	1 W03177	Rabbit poly-immunoglobulin	3.82e-06
11	209	7.4	584	1 R14671	Truncated poly Ig-rece	3.92e-06
12	167	5.9	205	1 Y05069	Human PIGR-2 protein s	3.54e-03
13	129	4.6	201	1 W99070	Human PIGR-1	1.23e+00
14	130	4.6	211	1 W86306	Kidney injury associat	1.06e+00
15	125	4.4	46	1 W99077	Immunoglobulin like pr	2.23e+00
16	123	4.3	195	1 W99071	Partial human PIGR-1	2.99e+00
17	122	4.3	303	1 R84103	Equine herpesvirus (EH	3.46e+00
18	122	4.3	303	1 R52667	Equine herpesvirus US2	3.46e+00
19	118	4.2	319	1 W14146	Human A33 antigen	6.19e+00
20	112	4.0	57	1 W99072	Immunoglobulin like pr	1.47e+01
21	112	4.0	60	1 W99073	Immunoglobulin like pr	1.47e+01
22	113	4.0	83	1 Y04776	Mycobacterium species	1.27e+01
23	113	4.0	104	1 Y04773	Mycobacterium species	1.27e+01

24	113	4.0	134	1	R04385	Colon Cancer monoclonal	1.27e+01
25	113	4.0	134	1	R38314	Sequence of murine ant	1.27e+01
26	113	4.0	139	1	W59615	Anti-RSV F protein Hul	2.25e+01
27	109	3.9	114	1	R48617	Sequence of the monocl	2.25e+01
28	110	3.9	138	1	W59617	Anti-RSV F protein Hul	1.95e+01
29	111	3.9	476	1	W01818	Primate anti-human	1.69e+01
30	111	3.9	476	1	W53761	Macaque primatized 7C1	1.69e+01
31	108	3.8	120	1	R50216	HSV glycoprotein F bin	2.59e+01
32	108	3.8	138	1	W59616	Anti-RSV F protein Hul	2.59e+01
33	107	3.8	138	1	R32246	BR55-2 murine IgG3 hea	2.98e+01
34	107	3.8	138	1	R32242	Chimeric MAB heavy cha	2.98e+01
35	107	3.8	138	1	W59618	Anti-RSV F protein Hul	2.98e+01
36	107	3.8	139	1	R31588	BR55-2 heavy chain var	2.98e+01
37	108	3.8	309	1	W83322	Single chain Apo-2 ant	2.59e+01
38	108	3.8	377	1	R53751	Seven transmembrane re	2.59e+01
39	108	3.8	377	1	W48732	Human R2 seven transme	2.59e+01
40	107	3.8	659	1	W11727	H-Delta-1 polypeptide	2.98e+01
41	106	3.7	97	1	R50100	Sequence encoded by hu	3.42e+01
42	106	3.7	118	1	R68739	MAB B3 heavy chain.	3.42e+01
43	106	3.7	247	1	R32091	H and L chain FV regio	3.42e+01
44	106	3.7	249	1	R95213	Heavy and light chain	3.42e+01
45	106	3.7	249	1	R32090	H and L chain FV regio	3.42e+01

ALIGNMENTS

RESULT 1  
ID Y05001 standard; Protein; 390 AA.  
AC Y05001;  
DT 16-JUN-1999 (first entry)  
DE Human PIGRL-1 protein sequence.  
KW PIGRL-1; human; autoimmune disease; hyper-IgM Immunodeficiency; HIM;  
KW X-linked Severe Combined Immunodeficiency; XSCID; IgA deficiency;  
KW diagnosis; therapy.  
OS Homo sapiens.  
PN EP-905238-A2.  
PD 31-MAR-1999.  
PF 14-AUG-1998; 306487.  
PR 30-OCT-1997; US-961564.  
PR 25-AUG-1997; US-056935.  
PI (SMIK ) SMITHLINE BEECHAM CORP.  
PA Sweet RW, Truneh A, Wu S;  
DR WPI; 99-192666/17.  
DR N-PSDB; X28178.  
PT New polypeptides encoding PIGRL-1 useful for treating diseases such  
PT as X-linked Severe Combined Immunodeficiency  
PS Claim 11; Page 7; 26pp; English.  
CC This sequence is the human PIGRL-1 protein of the invention.  
CC Autoimmune diseases involving altered expression or activity of PIGRL-1  
CC may include Hyper-IgM Immunodeficiency (HIM), X-linked Severe Combined  
CC Immunodeficiency (XSCID) and IgA deficiency. These diseases can be  
CC diagnosed or susceptibility to them predicted by: (1) determining whether  
CC there is a mutation in the genomic copy of the gene encoding PIGRL-1;  
CC (2) measuring the amount of PIGRL-1 in a sample derived from the patient.  
CC Patients deficient in PIGRL-1 can be treated by administering either the  
CC PIGRL-1 DNA or its complement or an agonist of PIGRL-1 to the patient.  
CC Patients with excessive expression or activity of PIGRL-1 can be treated  
CC by administering an antagonist of PIGRL-1, an antisense nucleic acid  
CC molecule which inhibits the expression of PIGRL-1 or administering  
CC sufficient PIGRL-1 to compete with the endogenous activity. PIGRL-1 can  
CC be used to identify its agonists by contacting a cell expressing PIGRL-1  
CC with a candidate compound in the presence of a signal system and noting  
CC the candidate as an agonist if a signal is produced. The same method can  
CC be used to identify antagonists of PIGRL-1 but the presence of an  
CC antagonist is indicated by a decrease in production of the signal.  
CC Antibodies against PIGRL-1 may be used to isolate or identify clones  
CC expressing PIGRL-1. Polynucleotides encoding PIGRL-1 may be used to  
CC identify chromosomal mutations in the gene encoding PIGRL-1 in patients.  
CC This information may then be correlated with the incidence of autoimmune  
CC disease in those patients to identify whether the mutation causes the  
CC disease.  
SQ Sequence 390 AA;

Query Match 99.5%; Score 2815; DB 1; Length 390;  
 Best Local Similarity 99.7%; Pred. No. 5,91e-218;  
 Matches 389; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 MDFWLPLYFLPYSGALRILPEVKVEGELGSGVTIKCPLEPMHVRVYLCREMAGSGTCGT 60  
 QY 1 MDRWLPLYFLPYSGALRILPEVKVEGELGSGVTIKCPLEPMHVRVYLCREMAGSGTCGT 60

Db 61 VYSTNFIKAEYKGRVTLKQYPRKNLFLVEVTQLTESDSGYACGAMTDRGKTQKYL 120  
 QY 61 VYSTNFIKAEYKGRVTLKQYPRKNLFLVEVTQLTESDSGYACGAMTDRGKTQKYL 120

Db 121 NVHSEYEPSWEOQMPETPKWFLPYLFQMPAYASSKFVTRVTTTAQRGKVPVPHHSSP 180  
 QY 121 NVHSEYEPSWEOQMPETPKWFLPYLFQMPAYASSKFVTRVTTTAQRGKVPVPHHSSP 180

Db 181 TQTLTHPRVSRASSVAGDKPRTFLPSTTASKISALEGLLKPTQPSYNHHTLHRQALD 240  
 QY 181 TQTLTHPRVSRASSVAGDKPRTFLPSTTASKISALEGLLKPTQPSYNHHTLHRQALD 240

Db 241 YGSGREGGPHILPTILGLFLALLGLVVKRAVERKALSRARRLAVMRALLESSQ 300  
 QY 241 YGSGREGGPHILPTILGLFLALLGLVVKRAVERKALSRARRLAVMRALLESSQ 300

Db 301 RPRGSPRPSQNNIYSACPRASCADAAGTGEAPVPGGAPLPAPLQVSSPWLHAPSL 360  
 QY 301 RPRGSPRPSQNNIYSACPRASCADAAGTGEAPVPGGAPLPAPLQVSSPWLHAPSL 360

Db 361 KTCSEYVSLYHOPAAAMDESDSDYINVPA 390  
 QY 361 KTCSEYVSLYHOPAAAMDESDSDYINVPA 390

## RESULT 2

ID Y05002 standard; peptide; 107 AA.  
 AC Y05002;  
 DE 16-JUN-1999 (first entry)  
 DE Human PIGRL-1 protein sequence fragment.  
 KW PIGRL-1; human; autoimmune disease; hyper-IgM Immunodeficiency; HIM;  
 KW X-linked Severe Combined Immunodeficiency; XSCID; IgA deficiency;  
 KW diagnosis; therapy.  
 OS Homo sapiens.  
 PN EP-905238-A2.  
 PD 31-MAR-1999.  
 PF 14-AUG-1998; 306487.  
 PR 30-OCT-1997; US-961564.  
 PR 25-AUG-1997; US-056935.  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 PI Sweet RW, Truneh A, Wu S;  
 DR WPI: 99-192666/17.  
 DR N-PSDB; X28179.  
 PT New polypeptides encoding PIGRL-1 useful for treating diseases such  
 PT as X-linked Severe Combined Immunodeficiency  
 PS Disclosure; Page 9; 26pp; English.  
 CC This sequence is a fragment of the human PIGRL-1 of the invention.  
 CC Autoimmune diseases involving altered expression or activity of PIGRL-1  
 CC may include Hyper-IgM Immunodeficiency (HIM), X-linked Severe Combined  
 CC Immunodeficiency (XSCID) and IgA deficiency. These diseases can be  
 CC diagnosed or susceptibility to them predicted by: (1) determining whether  
 CC there is a mutation in the genomic copy of the gene encoding PIGRL-1; or  
 CC (2) measuring the amount of PIGRL-1 in a sample derived from the patient.  
 CC Patients deficient in PIGRL-1 in a sample derived from the patient.  
 CC PIGRL-1 DNA or its complement or an agonist of PIGRL-1 to the patient.  
 CC Patients with excessive expression or activity of PIGRL-1 can be treated  
 CC by administering an antagonist of PIGRL-1, an antisense nucleic acid  
 CC molecule which inhibits the expression of PIGRL-1 or administering  
 CC sufficient PIGRL-1 to compete with the endogenous activity. PIGRL-1 can  
 CC be used to identify its agonists by contacting a cell expressing PIGRL-1  
 CC with a candidate compound in the presence of a signal system and noting  
 CC the candidate as an agonist if a signal is produced. The same method can  
 CC be used to identify antagonists of PIGRL-1 but the presence of an  
 CC antagonist is indicated by a decrease in production of the signal.  
 CC Antibodies against PIGRL-1 may be used to isolate or identify clones

CC expressing PIGRL-1. Polynucleotides encoding PIGRL-1 may be used to  
 CC identify chromosomal mutations in the gene encoding PIGRL-1 in patients.  
 CC This information may then be correlated with the incidence of autoimmune  
 CC disease in those patients to identify whether the mutation causes the  
 CC disease.  
 CC Sequence 107 AA;

Query Match 27.8%; Score 787; DB 1; Length 107;  
 Best Local Similarity 99.0%; Pred. No. 3.36e-51;  
 Matches 104; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 MDRWLPLYFLPYSGALRILPEVKVEGELGSGVTIKCPLEPMHVRVYLCREMAGSGTCGT 60  
 QY 1 MDRWLPLYFLPYSGALRILPEVKVEGELGSGVTIKCPLEPMHVRVYLCREMAGSGTCGT 60

Db 61 VYSTNFIKAEYKGRVTLKQYPRKNLFLVEVTQLTESDSGYACG 105  
 QY 61 VYSTNFIKAEYKGRVTLKQYPRKNLFLVEVTQLTESDSGYACG 105

## RESULT 3

ID W03180 standard; Protein; 771 AA.  
 AC W03180;  
 DE 24-FEB-1997 (first entry)  
 DE Mouse poly-immunoglobulin receptor.  
 KW Mouse; immunoglobulin; receptor; protection protein; mutans;  
 KW heavy chain; antigen binding domain; protection; pathogen;  
 KW mucosal; environment; gastrointestinal; passive; immunisation;  
 KW Guy's 13 antibody; prevention; dental caries; Streptococcus;  
 KW poly; sorbinus; murine.  
 OS Mus musculus.  
 PH Key  
 FT region 13..45  
 FT /note= "putative immunoglobulin binding residues  
 FT of domain I"  
 FT domain 1..120  
 FT /label= domain\_I  
 FT domain 110..230  
 FT /label= domain\_II  
 FT domain 210..340  
 FT /label= domain\_III  
 FT domain 320..450  
 FT /label= domain\_IV  
 FT domain 440..550  
 FT /label= domain\_V  
 FT region 550..606  
 FT /note= "external portions of domain VI"  
 FT region 550..627  
 FT /note= "external portions of domain VI"  
 FT region 625..660  
 FT /label= transmembrane\_segment  
 FT region 650..771  
 FT /label= intracellular\_portion  
 PN W09621012-A1.  
 PD 11-JUL-1996.  
 PF 27-DEC-1995; U16889.  
 PR 30-DEC-1994; US-367395.  
 PR 04-MAY-1995; US-434000.  
 PA (PLAN-) PLANT BIOTECHNOLOGY INC.  
 PA (UNNE-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.  
 PA (PLAN-) PLANET BIOTECHNOLOGY INC.  
 PI Hiatt AC, Lehner T, Ma JKC;  
 DR WPI: 96-333987/33.  
 DR N-PSDB; T31290.  
 PT Immunoglobulin and protection protein complex and its prodn. in  
 PT plants - useful for passive immunisation against mucosal antigens,  
 PT esp. against S. mutans and S. sorbinus to prevent dental caries  
 PS Disclosure; Pages 117-121; 152pp; English.  
 CC The present sequence is the mouse poly-immunoglobulin (Ig)  
 CC receptor, a portion of which corresp. to residues 1-627, pref.  
 CC 1-606, or esp. residues 13-45, 1-120, 110-230, 210-340, 320-450,  
 CC 440-550, 550-606 or 550-627 comprises a protection protein (PP).  
 CC The Ig of the invention comprises a PP as above in association with

CC an Ig derived heavy chain, having at least a portion of an antigen  
 CC (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.  
 CC gastrointestinal, environments, therefore enhancing its  
 CC effectiveness in passively immunising animals against mucosal  
 CC pathogens. The Ag binding domain is specifically derived from the  
 CC Guy's 13 antibody, and the Ig can be used to prevent dental caries  
 CC by binding, e.g. Streptococcus mutans serotypes c, e and f, or  
 CC S. sorbinus serotypes d and g.  
 SQ Sequence 771 AA;

Query Match 9.0%; Score 254; DB 1; Length 771;  
 Best Local Similarity 36.8%; Pred. No. 2.06e-09;  
 Matches 43; Conservative 26; Mismatches 42; Indels 6; Gaps 5;

Db 13 FSGVSTKPIGPQVEISGDSVITCYPPPTSVNRHTRKYWCRO-GASGMCTTLLSN 71  
 QY 10 FLFVSGALRIILPEVKVEGLGGSTIKC--P-LP-EHVRVYIYLCREWAGSGTCGVVST 65  
 Db 72 GYLKSEYSGRANLNFENNFTVINISQLODGTGSKCLG-TSNRGLSFDVSLV 127  
 QY 66 NFIRAKEYGRVTLKQYPRKNLFLVEVTQLTSDSGVYACGAGMNTDRGKTQKVTLLNV 122

## RESULT 4

ID W95601 standard; Protein; 607 AA.  
 AC W95601.  
 DT 08-JUN-1999 (first entry)  
 DE Human secretory immunoglobulin A component.  
 KW immunoglobulin A; secretory; component; IgA; human; treatment;  
 KW prevention; infection; HIV; AIDS; cold; flu; virus;  
 KW human immunodeficiency virus; respiratory syncytial virus.  
 OS Homo sapiens.  
 PN W09857993-A1.  
 PD 23-DEC-1998.  
 PF 10-JUN-1998; U11975.  
 PR 19-JUN-1997; US-050969.  
 PA (REGC ) UNIV CALIFORNIA.  
 PI Chintalacharuvu KR, Morrison SL;  
 DR WPI: 99-080950/07.  
 DR N-PSDB: X07407.  
 PT Producing secretory immunoglobulin in single cells - useful to  
 PT produce commercial quantities of secretory immunoglobulin to prevent  
 PT or treat infections  
 PS Disclosure; Pages 22-24; 39pp; English.  
 CC The sequence is that of the secretory component of human secretory  
 CC immunoglobulin A (sIgA). It can be used as part of a method for  
 CC the production of sig molecules. This method is useful for  
 CC producing commercial quantities of sig (especially sIgA) to treat  
 CC or prevent infections. In particular, sIgA produced by the method  
 CC can be used to prevent or treat infections in mammals, birds or  
 CC fish; especially systemic infections or infections at a mucosal  
 CC surface. It is especially useful to prevent or treat infection  
 CC with human immunodeficiency virus (HIV), respiratory syncytial  
 CC virus, flu virus or cold virus. The method allows production of  
 CC commercial quantities of sig molecules for therapeutic use, not  
 CC previously possible; production using non-plant cells and a  
 CC single cell type is more efficient than a previous multi-step  
 CC process of fusing recombinant plant cells, and avoids alterations  
 CC of the sig by plant cells. sIgA molecules are more stable  
 CC and resistant to proteolysis than previously used IgA molecules,  
 CC and can be administered to prevent as well as to treat infections,  
 CC unlike e.g. IgG and IgM molecules.  
 SQ Sequence 607 AA;

Query Match 8.9%; Score 252; DB 1; Length 607;  
 Best Local Similarity 35.0%; Pred. No. 2.90e-09;  
 Matches 41; Conservative 30; Mismatches 40; Indels 6; Gaps 4;

Db 13 FPAITKSPFGPEVNSVEGNSVITCYPPPTSVNRHTRKYWCRO-GARGGCITLISSE 71  
 QY 10 FLFVSGALRIILPEVKVEGLGGSTIKC--P-LP-EHVRVYIYLCREWAGSGTCGVVST 65  
 Db 72 GYVSSKYAGRANLTFNPENGFVYVNIQALSQDSDGRYKCGLGINS-RGLSFDVSLV 127

QY 66 NFIRAKEYGRVTLKQYPRKNLFLVEVTQLTSDSGVYACGAGMNTDRGKTQKVTLLNV 122

## RESULT 5

ID W03178 standard; Protein; 746 AA.  
 AC W03178;  
 DT 24-FEB-1997 (first entry)  
 DE Human poly-immunoglobulin receptor.  
 KW Human; immunoglobulin; receptor; protection protein; mutans;  
 KW heavy chain; antigen binding domain; protection; pathogen;  
 KW mucosal; environment; gastrointestinal; passive; immunisation;  
 KW Guy's 13 antibody; prevention; dental caries; Streptococcus;  
 KW poly; sorbinus.  
 OS Homo sapiens.  
 PN Key  
 FT region  
 FT 13..45 Location/Qualifiers  
 FT /note= "putative immunoglobulin binding residues  
 FT of domain I"  
 FT 1..120  
 FT /label= domain\_I  
 FT 110..230  
 FT /label= domain\_II  
 FT 210..340  
 FT /label= domain\_III  
 FT 320..450  
 FT /label= domain\_IV  
 FT 440..550  
 FT /label= domain\_V  
 FT 550..606  
 FT /note= "external portions of domain VI"  
 FT 550..627  
 FT /note= "external portions of domain VI"  
 FT 625..660  
 FT /label= transmembrane\_segment  
 FT 650..746  
 FT /label= intracellular\_portion  
 PN W09621012-A1.  
 PD 11-JUL-1996.  
 PF 27-DEC-1995; U16889.  
 PR 30-DEC-1994; US-367395.  
 PR 04-MAY-1995; US-434000.  
 PA (PLAN-) PLANT BIOTECHNOLOGY INC.  
 PA (UNME-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.  
 PA (PLAN-) PLANET BIOTECHNOLOGY INC.  
 PI Hiatt AC, Lehner T, Ma JKC;  
 DR WPI: 96-333987/33.  
 DR N-PSDB: T31288.  
 PT Immunoglobulin and protection protein complex and its prodn. in  
 PT plants - useful for passive immunisation against mucosal antigens,  
 PT esp. against S. mutans and S. sorbinus to prevent dental caries  
 PS Disclosure; Pages 105-108; 152pp; English.  
 CC The present sequence is the human poly-immunoglobulin (Ig)  
 CC receptor, a portion of which corresp. to residues 1-627, pref.  
 CC 1-606, or esp. residues 13-45, 1-120, 110-230, 210-340, 320-450,  
 CC 440-550, 550-606 or 550-627 comprises a protection protein (PP).  
 CC The Ig of the invention comprises a pp as above in association with  
 CC an Ig derived heavy chain, having at least a portion of an antigen  
 CC (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.  
 CC gastrointestinal, environments, therefore enhancing its  
 CC effectiveness in passively immunising animals against mucosal  
 CC pathogens. The Ag binding domain is specifically derived from the  
 CC Guy's 13 antibody, and the Ig can be used to prevent dental caries  
 CC by binding, e.g. Streptococcus mutans serotypes c, e and f, or  
 CC S. sorbinus serotypes d and g.  
 SQ Sequence 746 AA;

Query Match 8.7%; Score 247; DB 1; Length 746;  
 Best Local Similarity 37.7%; Pred. No. 6.77e-09;  
 Matches 40; Conservative 27; Mismatches 32; Indels 7; Gaps 5;  
 Db 7 PE-EVNSVEGNSVITCYPPPTSVNRHTRKYWCRO-GARGGCITLISSEGYVSSKYAGRA 64

QY 21 PEVKVEGELGSGVTIK---PLP-EMHVRVYLCREMAGSGTGTGVVSTTFNFIKAETKGRV 76  
 Db 65 NLTNFPENG: FVNIAQLSQDSGRYKCGLGINS-RGLSFDVSLV 109  
 QY 77 TLKQYPRKNFLVEVTQTESDSGVYACGAGNMDRGTKTKVILN 122  
 RESULT 6  
 AC W03179 standard; Protein; 757 AA.  
 DT 24-FEB-1997 (first entry)  
 DE Bovine poly-immunoglobulin receptor.  
 KW heavy chain; antigen binding domain; protection; pathogen;  
 KW mucosal; environment; gastrointestinal; passive; immunisation;  
 KW Guy's 13 antibody; prevention; dental caries; Streptococcus;  
 KW poly; sorbinus; cow.  
 OS Bos taurus.  
 FH Location/Qualifiers  
 FH Key 13..45  
 FH Region /note="putative immunoglobulin binding residues  
 FT of domain I"  
 FT domain 1..120  
 FT /label= domain\_I  
 FT domain 110..230  
 FT /label= domain\_II  
 FT domain 210..340  
 FT /label= domain\_III  
 FT domain 320..450  
 FT /label= domain\_IV  
 FT domain 440..550  
 FT /label= domain\_V  
 FT region 530..606  
 FT /note="external portions of domain VI"  
 FT region 550..627  
 FT /note="external portions of domain VI"  
 FT region 625..660  
 FT /label= transmembrane\_segment  
 FT region 650..757  
 FT /label= intracellular\_portion  
 PN W09621012-A1.  
 PD 11-JUL-1996.  
 PF 27-DEC-1995; U16889.  
 PR 30-DEC-1994; US-367395.  
 PR 04-MAY-1995; US-434000.  
 PA (PLAN-) PLANT BIOTECHNOLOGY INC.  
 PA (UNNE-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.  
 PI (PLAN-) PLANET BIOTECHNOLOGY INC.  
 PI Hiatt AC, Lehner T, Ma JKC;  
 DR WPI: 96-333987/33.  
 DR N-PSDB; T31289.  
 PT Immunoglobulin and protection protein complex and its prodn. in  
 PT esp. against S. mutans and S. sorbinus to prevent dental caries  
 PT Disclosure; Pages 111-113; 152pp; English.  
 CC The present sequence is the bovine poly-immunoglobulin (Ig)  
 CC receptor, a portion of which corresp. to residues 1-627, pref.  
 CC 1-606, or esp. residues 13-45, 1-120, 110-230, 210-340, 320-450,  
 CC 440-550, 550-606 or 550-627 comprises a protection protein (PP).  
 CC The Ig of the invention comprises a pp as above in association with  
 CC an Ig derived heavy chain, having at least a portion of an antigen  
 CC (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.  
 CC gastrointestinal, environments, therefore enhancing its  
 CC effectiveness in passively immunising animals against mucosal  
 CC pathogens. The Ag binding domain is specifically derived from the  
 CC Guy's 13 antibody, and the Ig can be used to prevent dental caries  
 CC by binding, e.g. Streptococcus mutans serotypes c, e and f, or  
 CC S. sorbinus serotypes d and g.  
 SQ Sequence 757 AA;  
 Query Match 8.7%; Score 245; DB 1; Length 757;  
 Best Local Similarity 36.7%; Pred. No. 9.49e-09;  
 Matches 40; Conservative 25; Mismatches 37; Indels 7; Gaps 5;

Db 25 PE-EVSGEGRSVSIKCYPTSVNRHRTKYWCRO-GAQCRTTLISSEGVSDYVGRA 82  
 QY 21 PEVKVEGELGSGVTIK---PLP-EMHVRVYLCREMAGSGTGTGVVSTTFNFIKAETKGRV 76  
 Db 83 NLTNFPESGTFVVDISHLTHKDSGRYKCGLGISS-RGLNFDVSLVSD 130  
 QY 77 TLKQYPRKNFLVEVTQTESDSGVYACGAGNMDRGTKTKVILNHSE 125  
 RESULT 7  
 ID W03181 standard; Protein; 769 AA.  
 AC W03181;  
 DT 24-FEB-1997 (first entry)  
 DE Rat poly-immunoglobulin receptor.  
 KW heavy chain; antigen binding domain; protection; pathogen;  
 KW mucosal; environment; gastrointestinal; passive; immunisation;  
 KW Guy's 13 antibody; prevention; dental caries; Streptococcus;  
 KW poly; sorbinus.  
 OS Rattus rattus.  
 FH Location/Qualifiers  
 FH Key 13..45  
 FH Region /note="putative immunoglobulin binding residues  
 FT of domain I"  
 FT domain 1..120  
 FT /label= domain\_I  
 FT domain 110..230  
 FT /label= domain\_II  
 FT domain 210..340  
 FT /label= domain\_III  
 FT domain 320..450  
 FT /label= domain\_IV  
 FT domain 440..550  
 FT /label= domain\_V  
 FT region 550..606  
 FT /note="external portions of domain VI"  
 FT region 550..627  
 FT /note="external portions of domain VI"  
 FT region 625..660  
 FT /label= transmembrane\_segment  
 FT region 650..769  
 FT /label= intracellular\_portion  
 PN W09621012-A1.  
 PD 11-JUL-1996.  
 PF 27-DEC-1995; U16889.  
 PR 30-DEC-1994; US-367395.  
 PR 04-MAY-1995; US-434000.  
 PA (PLAN-) PLANT BIOTECHNOLOGY INC.  
 PA (UNNE-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.  
 PI (PLAN-) PLANET BIOTECHNOLOGY INC.  
 PI Hiatt AC, Lehner T, Ma JKC;  
 DR WPI: 96-333987/33.  
 DR N-PSDB; T31291.  
 PT Immunoglobulin and protection protein complex and its prodn. in  
 PT plants - useful for passive immunisation against mucosal antigens,  
 PT esp. against S. mutans and S. sorbinus to prevent dental caries  
 PT Disclosure; Pages 123-127; 152pp; English.  
 CC The present sequence is the rat poly-immunoglobulin (Ig)  
 CC receptor, a portion of which corresp. to residues 1-627, pref.  
 CC 1-606, or esp. residues 13-45, 1-120, 110-230, 210-340, 320-450,  
 CC 440-550, 550-606 or 550-627 comprises a protection protein (PP).  
 CC The Ig of the invention comprises a pp as above in association with  
 CC an Ig derived heavy chain, having at least a portion of an antigen  
 CC (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.  
 CC gastrointestinal, environments, therefore enhancing its  
 CC effectiveness in passively immunising animals against mucosal  
 CC pathogens. The Ag binding domain is specifically derived from the  
 CC Guy's 13 antibody, and the Ig can be used to prevent dental caries  
 CC by binding, e.g. Streptococcus mutans serotypes c, e and f, or  
 CC S. sorbinus serotypes d and g.  
 SQ Sequence 769 AA;

Query Match 8.7%; Score 246; DB 1; Length 769;

Best Local Similarity 35.0%; Pred. No. 8.01e-09; Matches 41; Conservative 30; Mismatches 40; Indels 6; Gaps 5;

Db 13 FSGVSTQSPICPDQVSSIEGNSVINCYPDPVSVRHNRKWCRO-GANGYCATLISSN 71  
 QY 10 FLVPSGALRIPEVKVEGELGGVTKC--P-LP-EHVRIVLCREMAGSGTGGVWSTT 65  
 Db 72 GYLKEYSGRASLINFNPENSTFVNIHLQEDTGSYKCGLG-TTNRGLFFDVSLEY 127  
 QY 66 NFIAKEYGRVTKQYPRKNLFLVEVTLQTESDGSVYACGAGMNTDRGKTQKVLNV 122

#### RESULT 8

ID W50033 standard; Protein; 532 AA.  
 AC W50033; (first entry)  
 DE Human immunity related factor.  
 KW Lymph node; human; immunity related factor; research; treatment;  
 KW Immune disease; infectious disease.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Peptide 1..16  
 FT Peptide /label= sig\_peptide  
 FT Peptide 17..532  
 FT Peptide /label= mat\_peptide  
 PN J10072495-A.  
 PD 17-MAR-1998.  
 PR 11-JUN-1997; 153218.  
 PR 13-JUN-1996; JP-152362.  
 PA (ASAH) ASAH KASEI KOGYO KK.  
 DR WPI: 98-234766/21.  
 DR N-PSDB; V20383.  
 PT Immunity related factor - useful in the treatment of immune related  
 PT and infectious diseases  
 PS Claims 1, 3 and 4; pages 18-20; 21pp; Japanese.  
 CC The present sequence is a lymph node derived human immunity  
 CC related factor, which can be used to research and treat immune and  
 CC infectious diseases.  
 CC See also 532 AA.  
 SQ Sequence 532 AA;

Query Match 8.0%; Score 226; DB 1; Length 532;

Best Local Similarity 39.6%; Pred. No. 2.32e-07; Matches 36; Conservative 17; Mismatches 34; Indels 4; Gaps 3;

Db 70 VSEGEPCGAVTIOCHYAPSVNRHQRKWCRLGPPRWICQITIVSTNYTHRYRDRVALTD 129  
 QY 25 VEGELGGVTKCPL-PE-M--HVRIVLCREMAGSGTGGVWSTTNRKAEYKGRVTLKQ 80  
 Db 130 FQORGLFVVRSLQSPDDIGCYLIGISENN 160  
 QY 81 YPRKNLFLVEVTLQTESDGSVYACGAGMNTD 111

#### RESULT 9

ID R14670 standard; Protein; 584 AA.  
 AC R14670.  
 DE 30-JAN-1992 (first entry)  
 KW Truncated poly Ig-receptor encoded by allele no. 1.  
 KW Rabbit; insemination; pregnancy.  
 OS Oryctolagus cuniculus.  
 FH Key Location/Qualifiers  
 FT Peptide 1..18  
 FT domain 10..118  
 FT domain /label= signal sequence  
 FT domain /number= I  
 FT domain /note= "poly-Ig binding"  
 FT domain 119..223  
 FT domain /number= II  
 FT domain 224..332  
 FT domain /number= III  
 FT domain 333..441  
 FT domain /number= IV

FT domain 442..552  
 FT /number= V  
 FT domain 553..584  
 FT /number= IV  
 FT /note= "incomplete"  
 PN W0116061-A.  
 PD 31-OCT-1991.  
 PF 16-APR-1991; U02604.  
 PR 16-APR-1990; US-510161.  
 PA (HARD) HARVARD COLLEGE.  
 PA (SURE-) INST SUISSE RECH EXP C.  
 PI Kraehenbuhl JP, Weltzin RA, Neutra MR;  
 DR WPI: 91-339549/46.  
 DR N-PSDB; Q14498.  
 PT Stabilised poly-Ig complex contg. portion of poly-Ig receptor -  
 PT useful in protection against pathogens or against pregnancy  
 PS Disclosure; Fig 7; 51 pp; English.  
 CC The sequence was deduced from a cDNA clone of allele no. 1 and  
 CC is a truncated poly-Ig receptor. The native gene (Mostov et al)  
 CC is mutated to delete the portion encoding the transmembrane and  
 CC intra-cellular domains. The recombinant protein produced by a  
 CC expression of the sequence is used as a stabiliser protein with a  
 CC compsn. can be administered directly to the mucosal surfaces of a  
 CC mammal to protect against a pathogen or against insemination. It  
 CC protects against allergens that contact the respiratory or digestive  
 CC mucosal surfaces and protects against pregnancy by cross-linking  
 CC sperm in the vagina.  
 CC See also R14671.  
 SQ Sequence 584 AA;

Query Match 7.9%; Score 223; DB 1; Length 584;

Best Local Similarity 45.5%; Pred. No. 3.83e-07; Matches 40; Conservative 11; Mismatches 30; Indels 7; Gaps 5;

Db 39 GDSVSICTYPTTSVTRHSRKFWCRE-EESGRCVTLAST-GYTSEYSGRGLTDFPDKG 96  
 QY 30 GGSVTKC--PLPEM--HVRIVLCREMAGSGTGGVWSTTNRKAEYKGRVTLKQYPRKN 85  
 Db 97 EFVTVLDLTQNDSSYKCGVGN-GRG 123  
 QY 86 LFLVEVTLQTESDGSVYACGAGMNTDRG 113

#### RESULT 10

ID W03177 standard; Protein; 773 AA.  
 AC W03177;  
 DT 24-FEB-1997 (first entry)  
 DE Rabbit poly-immunoglobulin receptor.  
 KW heavy chain; immunoglobulin; receptor; protection protein; mutants;  
 KW mucosal; environment; gastrointestinal; passive; immunisation;  
 KW Guy's 13 antibody; prevention; dental caries; Streptococcus;  
 KW poly; sorbinus.  
 OS Oryctolagus cuniculus.  
 FH Key Location/Qualifiers  
 FT region 21..43  
 FT /note= "immunoglobulin binding residues of domain I"  
 FT domain 1..118  
 FT /label= domain\_I  
 FT domain 119..223  
 FT /label= domain\_II  
 FT domain 224..332  
 FT /label= domain\_III  
 FT domain 333..441  
 FT /label= domain\_IV  
 FT domain 442..552  
 FT /label= domain\_V  
 FT region 553..606  
 FT /note= "external portions of domain VI"  
 FT region 553..627  
 FT /note= "external portions of domain VI"

FT region 630..652  
 FT /label= transmembrane\_segment  
 FT 653..755  
 FT /label= intracellular\_portion  
 PN W09621012-A1.  
 PD 11-JUL-1996.  
 PF 27-DEC-1995; US-367395.  
 PR 30-DEC-1994; US-434000.  
 PA (PLAN-) PLANT BIOTECHNOLOGY INC.  
 PA (UNNE-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.  
 PA (PLAN-) PLANET BIOTECHNOLOGY INC.  
 PI Blatt AC, Lehner T, Ma JKC;  
 DR WPI: 96-333987/33.  
 DR N-PSDB: T31287.  
 PT Immunoglobulin and protection protein complex and its prodn. in  
 PT plants - useful for passive immunisation against mucosal antigens,  
 PT esp. against S. mutans and S. sorbinus to prevent dental caries  
 FS Claim 10; Pages 98-102; 152pp; English.  
 CC The present sequence is the rabbit poly-immunoglobulin (Ig)  
 CC receptor, a portion of which corresp. to residues 1-627, pref.  
 CC 1-606, or esp. residues 21-43, 1-118, 119-223, 224-332, 333-441,  
 CC 442-552, 553-606 or 553-627 comprises a protection protein (PP).  
 CC The Ig of the invention comprises a PP as above in association with  
 CC an Ig derived heavy chain having at least a portion of an antigen  
 CC (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.  
 CC gastrointestinal, environments, therefore enhancing its  
 CC effectiveness in passively immunising animals against mucosal  
 CC pathogens. The Ag binding domain is specifically derived from the  
 CC Guy's 13 antibody, and the Ig can be used to prevent dental caries  
 CC by binding, e.g. Streptococcus mutans serotypes c, e and f, or  
 CC S. sorbinus serotypes d and g.  
 SQ Sequence 773 AA;  
 Query Match 7.9%; Score 223; DB 1; Length 773;  
 Best Local Similarity 45.5%; Pred. No. 3.83e-07;  
 Matches 40; Conservative 11; Mismatches 30; Indels 7; Gaps 5;  
 Db 39 GDSVITCYPTTSVNRHSRKFRCRE-EESGRGVTLAST-GYTSQYSGRGLDFFDPRG 96  
 Qy 30 GGSVTIKC--PLPEN--HVRIYLCREMGAGSGCTGVVSTTFIKAEYKGRVTLKQYPRN 85  
 Db 97 EFVVTVDLTNDSDSGYKCGVGN-GRG 123  
 Qy 86 LFLVEVTQLTESDSGYACGAGMNTDRG 113  
 RESULT 11  
 ID R14671 standard; Protein; 584 AA.  
 AC R14671.  
 DT 30-JAN-1992 (first entry)  
 DE Truncated poly Ig-receptor encoded by allele no. 2.  
 KW Rabbit; insemination; pregnancy.  
 OS Oryctolagus cuniculus.  
 FH Key Location/Qualifiers  
 FT peptide 1..118  
 FT domain 10..118  
 FT /label= signal sequence  
 FT /number= I  
 FT domain 119..223  
 FT /number= II  
 FT domain 224..332  
 FT /number= III  
 FT domain 333..441  
 FT /number= IV  
 FT domain 442..552  
 FT /number= V  
 FT domain 553..584  
 FT /number= IV  
 FT /note= "incomplete"  
 PN W09116061-A.  
 PD 31-OCT-1991.

PF 16-APR-1991; U02604.  
 PR 16-APR-1990; US-510161.  
 PA (HARD ) HARVARD COLLEGE.  
 PA (SURE-) INST SUISSE RECH EXPER C.  
 PI Kraehenbuhl JP, Weltzin RA, Neutra MR;  
 DR WPI: 91-339549/46.  
 DR N-PSDB: Q14499.  
 PT Stabilised poly-Ig complex contg. portion of poly-Ig receptor -  
 PT useful in protection against pathogens or against pregnancy  
 PS Disclosure: Fig. 7: 51 pp; English.  
 CC The sequence was deduced from a cDNA clone of allele no. 2 and  
 CC is a truncated poly-Ig receptor. The native gene (Mostov et al)  
 CC is mutated to delete the portion encoding the transmembrane and  
 CC intra-cellular domains. The recombinant protein produced by  
 CC expression of the sequence is used as a stabiliser protein with a  
 CC poly-Ig specific for a selected antigen or family of antigens. The  
 CC compsn. can be administered directly to the mucosal surfaces of a  
 CC mammal to protect against a pathogen or against insemination. It  
 CC protects against allergens that contact the respiratory or digestive  
 CC mucosal surfaces and protects against pregnancy by cross-linking  
 CC sperm in the vagina.  
 CC See also R14670.  
 SQ Sequence 584 AA;  
 Query Match 7.4%; Score 209; DB 1; Length 584;  
 Best Local Similarity 43.2%; Pred. No. 3.32e-06;  
 Matches 38; Conservative 13; Mismatches 30; Indels 7; Gaps 5;  
 Db 39 GDSVITCYPTTSVNRHSRKFRCRE-EESGRGVTLAST-GYTSQYSGRGLDFFDPRG 96  
 Qy 30 GGSVTIKC--PLPEN--HVRIYLCREMGAGSGCTGVVSTTFIKAEYKGRVTLKQYPRN 85  
 Db 97 EFVVTVDLTNDSDSGYKCGVGN-GRG 123  
 Qy 86 LFLVEVTQLTESDSGYACGAGMNTDRG 113  
 RESULT 12  
 ID Y05069 standard; Protein; 205 AA.  
 AC Y05069.  
 DT 16-JUN-1999 (first entry)  
 DE Human PIGR-2 protein sequence.  
 KW PIGR-2; human; autoimmune disease; rheumatoid arthritis; psoriasis;  
 KW Multiple Sclerosis; Systemic Lupus Erythematosus; diagnosis; therapy;  
 KW Inflammatory Bowel Disease.  
 OS Homo sapiens.  
 PN EP-905237-A2.  
 PD 31-MAR-1999.  
 PF 07-AUG-1998; 306323.  
 PR 21-NOV-1997; US-976293.  
 PR 25-AUG-1997; US-056774.  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 PI Sweet RW, Truneh A, Wu S;  
 DR WPI: 99-192665/17.  
 DR N-PSDB: X28250.  
 PT New polypeptides encoding human PIGR-2 useful for treating diseases  
 PT such as rheumatoid arthritis and multiple sclerosis  
 PS Claim 11; Page 17; 23pp; English.  
 CC This sequence is the human PIGR-2 protein of the invention.  
 CC Autoimmune diseases involving altered expression or activity of PIGR-2  
 CC may include rheumatoid arthritis, Multiple Sclerosis, psoriasis,  
 CC Systemic Lupus Erythematosus and Inflammatory Bowel Disease. These  
 CC diseases can be diagnosed or susceptibility to them predicted by:  
 CC (1) determining whether there is a mutation in the genomic copy of the  
 CC gene encoding PIGR-2; or (2) measuring the amount of PIGR-2 in a sample  
 CC derived from the patient. Patients deficient in PIGR-2 can be treated by  
 CC administering either the PIGR-2 DNA or its complement or an agonist of  
 CC PIGR-2 to the patient. Patients with excessive expression or activity of  
 CC PIGR-2 can be treated by administering an antagonist of PIGR-2, an  
 CC antisense nucleic acid molecule which inhibits the expression of PIGR-2  
 CC or administering sufficient PIGR-2 to compete with the endogenous  
 CC activity. PIGR-2 can be used to identify its agonists by contacting a  
 CC cell expressing PIGR-2 with a candidate compound in the presence of a







OM of: US-09-135-238B-2 to: N\_Geneseq\_36.\* out\_format : pfs  
Date: Sep 12, 2000 5:36 AM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlp  
-O=/cgn2\_1/USPTO\_spool/US09135238/runat\_24082000\_091114\_16813/app\_query.fasta\_1.1282  
-DB=N\_Geneseq\_36 -OFMT=fastap -SUFFIX=p2n.rng -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.500  
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-GAPOP=6.000 -GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALLGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM-ext -MINLEN=0 -MAXLEN=1000000  
-USER=US09135238 @CGN1\_1\_123 -NCPU=6 -ICPU=3 -LONGLOG -NO\_XLPHY  
-WAIT -THREADS=1

Search information block:  
Query: US-09-135-238B-2  
Query length: 390  
Database: N\_Geneseq\_36.\*  
Database sequences: 311585  
Database length: 125096042  
Search time (sec): 122.100000

score_list:	Sequence	Strd Orig	zScore	EScore	Len	Documentation
	N_Geneseq_36:X28178	+ 2047.00	2129.34	5.3e-111	2040	Human PIGRL-1 coding sequence.
	N_Geneseq_36:X28179	+ 1125.00	1174.09	8.5e-58	1047	Human PIGRL-1 coding sequence.
	N_Geneseq_36:T31290	+ 198.00	199.21	0.0017	3095	Mouse poly-immunoglobulin rece
	N_Geneseq_36:T31291	+ 188.00	170.27	0.0697	29392	Mouse poly Ig receptor protei
	N_Geneseq_36:T31292	+ 187.50	192.24	0.0042	1911	cDNA for human immunity relat
	N_Geneseq_36:T31293	+ 185.50	185.74	0.0096	3269	Rat poly-immunoglobulin recept
	N_Geneseq_36:T31294	+ 185.00	189.53	0.0059	1936	Secretory component DNA fragme
	N_Geneseq_36:T31295	+ 182.00	186.82	0.0083	1839	Human secretory immunoglobulin
	N_Geneseq_36:T31296	+ 182.00	186.01	0.0093	2031	Secretory component DNA fragme
	N_Geneseq_36:T31297	+ 182.00	183.23	0.0136	2919	Human poly-immunoglobulin rece
	N_Geneseq_36:T31298	+ 182.00	191.02	0.0171	3630	Bovine poly-immunoglobulin rece
	N_Geneseq_36:T31299	+ 171.00	174.33	0.0414	2084	Rat kidney injury related mole
	N_Geneseq_36:T31300	+ 171.00	172.62	0.0515	2586	Rat kidney injury related mole
	N_Geneseq_36:T31301	+ 166.50	170.51	0.0676	1876	Rat kidney injury related mole
	N_Geneseq_36:T31302	+ 166.50	165.34	0.1312	3517	Rabbit poly-immunoglobulin rece
	N_Geneseq_36:T31303	+ 159.50	167.34	0.1015	1137	Human PIGRL-2 coding sequence.
	N_Geneseq_36:T31304	+ 145.00	148.11	1.20	1876	Clone 1 for truncated poly Ig-
	N_Geneseq_36:T31305	+ 143.50	133.93	7.37	8690	Plasmid pD17-hJm14.H1. Inhibit
	N_Geneseq_36:T31306	+ 142.00	130.90	0.8361	914	Human EST R2810 DNA. Complexes
	N_Geneseq_36:T31307	+ 139.50	130.12	12.02	8321	Plasmid expressing hBR96-A. In
	N_Geneseq_36:T31308	+ 138.50	146.63	1.44	985	Mycobacterium tuberculosis anti
	N_Geneseq_36:T31309	+ 138.50	146.63	1.44	985	M. tuberculosis immunogenic pol
	N_Geneseq_36:T31310	+ 136.50	149.12	1.05	565	EST clone EN260. New polynucle
	N_Geneseq_36:T31311	+ 136.50	140.14	3.32	1683	Human dUTPase gene 5' region.
	N_Geneseq_36:T31312	+ 136.00	123.79	27.06	11529	Plasmid pHA4602. Antibody con
	N_Geneseq_36:T31313	+ 135.00	123.35	28.60	10704	Plasmid pHA4625. Antibody con
	N_Geneseq_36:T31314	+ 131.00	118.16	55.71	12132	Plasmid pHA4807. Antibody con
	N_Geneseq_36:T31315	+ 130.00	133.70	7.59	1617	Antibody D heavy chain. Prodn.
	N_Geneseq_36:T31316	+ 129.50	134.18	7.13	1431	Primate anti-human B7.1 ant
	N_Geneseq_36:T31317	+ 129.50	134.18	7.13	1431	Macaque primatized 7C10 heavy
	N_Geneseq_36:T31318	+ 127.00	128.05	15.66	2197	IS1096 transposon. Leucine aux
	N_Geneseq_36:T31319	+ 126.50	120.88	39.26	4926	Plasmid pHC665 nucleotide sequ
	N_Geneseq_36:T31320	+ 126.50	114.03	94.63	11336	Approximate nucleotide sequen
	N_Geneseq_36:T31321	+ 125.50	129.22	13.48	1576	Human anti-HBs heavy chain. Hu
	N_Geneseq_36:T31322	+ 125.50	129.22	13.48	1576	Human anti-HBs heavy chain. Hu
	N_Geneseq_36:T31323	+ 125.50	122.89	30.35	3400	3F4 human G2/G4 chimeric antib
	N_Geneseq_36:T31324	+ 125.50	116.64	67.64	7266	Human calcium channel a1B subu
	N_Geneseq_36:T31325	+ 124.50	116.45	69.37	6557	Anti-IgE VH expression vector.
	N_Geneseq_36:T31326	+ 124.50	108.70	187.39	16812	HSV-2 strain S85 Contig ID 12
	N_Geneseq_36:T31327	+ 124.00	125.90	20.63	1951	C. acidivorans gamma-lactamase
	N_Geneseq_36:T31328	+ 124.00	124.32	25.27	2364	CD4-specific CDR-grafted heavy

N\_Geneseq\_36:T50962 + 124.00 115.30 80.34 7073 ! TF8-5G9 CDR-grafted heavy c  
N\_Geneseq\_36:T57472 + 123.50 125.47 21.80 1929 ! Sorghum bicolor (L.) Moench  
N\_Geneseq\_36:T96035 + 123.00 125.54 21.60 1795 ! Human kidney injury related  
N\_Geneseq\_36:T72885 + 123.00 113.19 105.27 8051 ! Sugar biosynthesis gene clu  
N\_Geneseq\_36:X25774 + 123.00 113.08 106.78 8160 ! S.erythraea erythromycin-sy

seq\_name: N\_Geneseq\_36:X28178

seq\_documentation\_block:

ID X28178 standard: cDNA; 2040 BP.

AC X28178;

DE 16-JUN-1999 (first entry)

DT Human PIGRL-1 coding sequence.

KW PIGRL-1; human; autoimmune disease; hyper-IgM Immunodeficiency; HIM;

KW X-linked Severe Combined Immunodeficiency; XSCID; IgA deficiency;

KW diagnosis; therapy; ss.

OS Homo sapiens.

PN EP-905238-A2.

PF 31-MAR-1999.

PF 14-AUG-1998; 306487.

PR 30-OCT-1997; US-961564.

PR 25-AUG-1997; US-056935.

PA (SMIK ) SMITHKLINE BEECHAM CORP.

PI Sweet RW, Truneh A, Wu S;

DR WPI: 99-192666/17.

DR P-PSDB: Y05001.

PT New polypeptides encoding PIGRL-1 useful for treating diseases such

PT as X-linked Severe Combined Immunodeficiency

PS Claim 4; Page 18-19; 26pp; English.

CC This sequence encodes the human PIGRL-1 protein of the invention.

CC Autoimmune diseases involving altered expression or activity of PIGRL-1

CC may include Hyper-IgM Immunodeficiency (HIM), X-linked Severe Combined

CC Immunodeficiency (XSCID) and IgA deficiency. These diseases can be

CC diagnosed or susceptibility to them predicted by: (1) determining whether

CC there is a mutation in the genomic copy of the gene encoding PIGRL-1; or

CC (2) measuring the amount of PIGRL-1 in a sample derived from the patient.

CC Patients deficient in PIGRL-1 can be treated by administering either the

CC PIGRL-1 DNA or its complement or an agonist of PIGRL-1 to the patient.

CC Patients with excessive expression or activity of PIGRL-1 can be treated

CC by administering an antagonist of PIGRL-1, an antisense nucleic acid

CC molecule which inhibits the expression of PIGRL-1 or administering

CC sufficient PIGRL-1 to compete with the endogenous activity. PIGRL-1 can

CC be used to identify its agonists by contacting a cell expressing PIGRL-1

CC with a candidate compound in the presence of a signal system and noting

CC the candidate as an agonist if a signal is produced. The same method can

CC be used to identify antagonists of PIGRL-1 but the presence of an

CC antagonist is indicated by a decrease in production of the signal.

CC Antibodies against PIGRL-1 may be used to isolate or identify clones

CC expressing PIGRL-1. Polynucleotides encoding PIGRL-1 may be used to

CC identify chromosomal mutations in the gene encoding PIGRL-1 in patients.

CC This information may then be correlated with the incidence of autoimmune

CC disease in those patients to identify whether the mutation causes the

CC disease. 2040 BP; 478 A; 616 C; 494 G; 452 T;

SQ Sequence

alignment\_scores:

Quality: 2047.00 Length: 390

Ratio: 5.262 Gaps: 0

Percent Similarity: 99.744 Percent Identity: 99.744

alignment\_block:

US-09-135-238B-2 x X28178 ..

Align seg 1/1 to: X28178 from: 1 to: 2040

1 MetaspargtrpIeuTrpProLeuTyrPheLeuProValSerGlyAlaLe 17

||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

146 ATGGACTTCCTGGCTTTGGCCACCTTACTTCTCGCAGTATCAGGGGCCCT 195

||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

17 uatqIleleuProGluValGluGlyGluLeuGlyGlySerValt 34

||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

196 GAGGATCTCCAGAGTAAGGTAGAGGGGAGCTGGGGGATCAGTTA 245

```

34  hrIleLysCysProLeuProGluMetHisValArgIleTyrLeuCysArg 50
|||||
246 CCATCAAGTGCCTCTCTGAAATGCATGTGAGATATATCTGTGCGG 295
51  GluMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsp 67
|||||
296 GAGATGGCTGGATCTGGAACATGTGTACCTACCTGGTATCCACCACTT 345
67  eIleLysAlaGlyTyrLysGlyArgValThrLeuLysGlnTyrProArgL 84
|||||
346 CATCAGGCGAGAAATACAGGCGGAGTTACTCTGAGCAATACCCAGCA 395
84  ysAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerGly 100
|||||
396 AGAATCTGTCTTCTAGTGGAGGTAAACACAGCTGACAGAAAGTGACAGCGA 445
101 ValTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLy 117
|||||
446 GTCTATGCTCGGAGCGGCATGAACACAGACCGGGGAAAGACCCAGAA 495
117 sValThrLeuAsnValHisSerGluTyrGluProSerTrpGluGlnP 134
|||||
496 AGTCACCTGTAATCTCCACAGTGAATACGAGCCATCATGGGAAGCAGC 545
134 roMetProGluThrProLysTrpPheHisLeuProTyrLeuPheGluMet 150
|||||
546 CAATGCCTGAGACTCCAAATGGTTTCATCTGCCCTATTGTTCAGATG 595
151 ProAlaTyrAlaSerSerLysPheValThrArgValThrThrProAl 167
|||||
596 CTGTCATATGCCAGTCTTCCAAATTCGTACACAGAGTTACCCACAGC 645
167 aGlnArgGlyLysValProProValHisHisSerSerProThrThrGlnI 184
|||||
646 TCMAAGGGGAGAGTCTCTCCAGTTCACCACTCTCCCCACACCCAAA 695
184 leThrHisArgProArgValSerArgAlaSerSerValAlaGlyAspLys 200
|||||
696 TCACCCAGCGCCCTCGAGTGTCCAGAGCATCTTCAGTAGCAGGTGACAAG 745
201 ProArgThrPheLeuProSerThrThrAlaSerLysIleSerAlaLeuG 217
|||||
746 CCGCGAAGCTTCTGCGATCCACTACAGCTTCAAAATCTCAGTCTCGGA 795
217 uGlyLeuLeuLysProGlnThrProSerTyrAsnHisHisThrArgLeuH 234
|||||
796 GGGGCTGTCTAAGCCCGACAGCCAGCTACAAACACACACAGGCTGC 845
234 isArgGlnArgAlaLeuAspTyrGlySerGlnSerGlyArgGluGln 250
|||||
846 ACAGCAGAGAGCACTGGACTATGGCTACAGTCTGGAGGAAAGGCCAA 895
251 GlyPheHisIleLeuLeuProThrIleLeuGlyLeuPheLeuLeuAlaLe 267
|||||
896 GGATTTCAATCTGATCCCGACCATCTGGGCTTTTCTGCTGGCACT 945
267 uLeuGlyLeuValValLysArgAlaValGluArgArgLysAlaLeuSerA 284
|||||
946 TCTGGGGCTGGTGTGAAAGGGCGGTGAAAGGGGAAAGCCCTCTCCA 995
284 rArgAlaArgArgLeuAlaValArgMetArgAlaLeuGluSerSerGln 300
|||||
996 GCGCGGCCCGCGCTGCGCGTGAGGATGCGCGCTTCGGAGAGCTCCAG 1045
301 ArgProArgGlySerProArgProArgSerGlnAsnAsnIleTyrSerAl 317
|||||
1046 AGGCCCCCGGGTGGCGCGAGCGCGCTCCCAAAACAAACATCTACAGCG 1095
317 aCysProArgAlaArgGlyAlaAspAlaAlaGlyThrGlyGluAlap 334
|||||
1096 CTGCCCCGGCGCTCTGTGGAGGAGCGCTGCAGGCACAGGGGAGGCC 1145

```

```

334 toValProGlyProGlyAlaProLeuProAlaProLeuGlnValSer 350
|||||
1146 CGGTTCGCCGCCCGGAGCGCGTGGCCCCCGCCCGCTCGAGGTCTCT 1195
351 GluSerProTrpLeuHisAlaProSerLeuLysThrSerCysGluTyrVa 367
|||||
1196 GAATCTCCTCGCTCATGCGCCCATCTCTGAAGACGAGCTGTGAATACGT 1245
367 lSerLeuTyrHisGlnProAlaAlaMetMetGluAspSerAspSerAspA 384
|||||
1246 GAGCCTCTACCAACAGCTCGCCCATGATGGAGGACAGTGTATTCAGATG 1295
384 spTyrIleAsnValProAla 390
|||||
1296 ACTACATCAATGTTCTGCC 1315

```

seq\_name: N\_Geneseq\_36.X28179

seq\_documentation\_block:

ID X28179 standard; cDNA; 1047 BP.

AC X28179;

DT 16-JUN-1999 (first entry)

DE Human FIGRL-1 coding sequence fragment.

KW FIGRL-1; human; autoimmune disease; hyper-IgM Immunodeficiency; HIM;

KW X-linked Severe Combined Immunodeficiency; XSCID; IgA deficiency;

KW diagnosis; therapy; ss.

OS Homo sapiens.

PN EP-905238-A2.

PD 31-MAR-1999.

PF 14-AUG-1998; 306487.

PR 30-OCT-1997; US-961564.

PR 25-AUG-1997; US-056935.

PA (SWIK ) SMITHKLINE BEECHAM CORP.

PI Sweet RW, Truneh A, Wu S;

DR WPI; 99-192666/17.

DR P-PSDB; Y05002.

PT New polypeptides encoding FIGRL-1 useful for treating diseases such

as X-linked Severe Combined Immunodeficiency

PS Disclosure; Page 8; 26pp; English.

CC This sequence encodes the human FIGRL-1 protein of the invention.

CC Autoimmune diseases involving altered expression or activity of FIGRL-1

may include hyper-IgM Immunodeficiency (HIM), X-linked Severe Combined

Immunodeficiency (XSCID) and IgA deficiency. These diseases can be

diagnosed or susceptibility to them predicted by: (1) determining whether

there is a mutation in the genomic copy of the gene encoding FIGRL-1; or

(2) measuring the amount of FIGRL-1 in a sample derived from the patient.

CC Patients deficient in FIGRL-1 can be treated by administering either the

FIGRL-1 DNA or its complement or an agonist of FIGRL-1 to the patient.

CC Patients with excessive expression or activity of FIGRL-1 can be treated

by administering an antagonist of FIGRL-1, an antisense nucleic acid

molecule which inhibits the expression of FIGRL-1 or administering

sufficient FIGRL-1 to compete with the endogenous activity. FIGRL-1 can

be used to identify its agonists by contacting a cell expressing FIGRL-1

with a candidate compound in the presence of a signal system and noting

the candidate as an agonist if a signal is produced. The same method can

be used to identify antagonists of FIGRL-1 but the presence of an

antagonist is indicated by a decrease in production of the signal.

CC Antibodies against FIGRL-1 may be used to isolate or identify clones

expressing FIGRL-1. Polynucleotides encoding FIGRL-1 may be used to

identify chromosomal mutations in the gene encoding FIGRL-1 in patients.

CC This information may then be correlated with the incidence of autoimmune

disease in those patients to identify whether the mutation causes the

CC disease.

SQ Sequence 1047 BP; 255 A; 281 C; 276 G; 223 T;

alignment\_scores:

Quality: 1125.00 Length: 275

Ratio: 4.429 Gaps: 12

Percent Similarity: 92.364 Percent Identity: 90.182

alignment\_block:

US-09-135-238B-2 x X28179

Align seg 1/1 to: X28179 from: 1 to: 1047

```

1 MetAspArgTyrPheLeuTyrPheLeuProValSerGlyAlaLe 17
150 ATGGACTTCTGGCTTGGCCACTTACTTCTGCGCAGTATCAGGGCCCT 199
17 uArgileLeuProGluValValGlyValGluGlyGlySerValT 34
200 GAGGATCTCCAGAGCTAAGGTAGAGGGGAGCTGGCGGATCAGTTA 249
34 hrIleLysCysProLeuProGluMetHisValArgIleTyrLeuCysArg 50
250 CCATCAAGTGCACCTCTCTGAAATGCAATGTGAGGATATATCTGTGCGG 299
51 GluMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsnPh 67
300 GAGTGGCTGGATCTGGACATGTGGTACCTGGTATCCACCACTT 349
67 eIleLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProArgL 84
350 CATCAAGGCAGAAATCAAGGGCCGAGTTACTCTGAAGCAATACCCAGCA 399
84 ysAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerGly 100
400 AGAATCTGTCTCTAGTGGAGGTAACACAGCTGACAGAAAGTGACGGGA 449
101 ValTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLy 117
450 GTCTATGCTCGGGAGC.GGCATGAACACAGACCCGGGAAAGACCCAGAA 498
117 sValThrLeuAsnValHisSerGlyTyrGluProSerTrpGluGlnP 134
499 AGTCACCTTGATGCCAGAGTAATACGACCATCATGGAGAGAGC 548
134 roMetProGluThrProLysTrpPheHisLeuProTyrLeuPheGlnMet 150
549 CAATGCTCAGACTCCAAATGTTTCTATCTGCCTATTTGTTCCAGATG 598
151 ProAlaTyrAlaSerSerSerLysPheValThr...ArgValThrPro 166
599 CTGCAATATGCGGCTTCTCCACATTCGTAAACCCGAGAGTTACACACCA 648
167 Ala.GlnArgGlyLysValProValHisHisSerSerProThrThrG 183
649 GCTTCAAGGGGCAAGTCCCTCCAGTTCCACACTCTCTCCCCACACC 698
183 lnIle.ThrHisArgPro.ArgValSerArgAlaSerSerValAlaGlyA 199
699 AAATTCACCCACCGCCCTTCGAGTGTNCAGAGCATCTTCAGTAGCAGGTG 748
199 sPlysProArg.ThrPheLeuProSerThrThrAlaSerLysIleSerAl 215
749 ACAAGCCCGAAATTTCTGTCATCCACTACAGCTCAAAATCTCAGC 798
215 aLeuGluGlyLeuLeuLys...ProGlnThrProSerTyrAsnHisIst 231
799 TCTGGAAGGCGTCTTCAAGCCCAAGAGCCGCCAGCTACAAACAN.CACA 847
231 hrArgLeuHisArgGlnArgAlaLeuAsp...Tyr.GlySerGlnSerGl 246
848 CCAGGCTGCACAGGACAGACACTGGATCTATATGGGNTCACAGTCTGG 897
246 yArgGluGly.....GlnGlyPhe...HisIleLeuIlePro.ThrIle 259
898 G...GAGGGAANGNCAAGGATTTTNCATCTCTCTGATTCCTCCGACCATC 944
260 LeuGlyLeuPhe 263
945 NTGCGGCTTT 956

```

seq\_name: N\_Geneseq\_36.T31290

seq\_documentation\_block:

```

ID T31290 standard; CDNA; 3095 BP.
AC T31290;
DE 24-FEB-1997 (first entry)
KW Mouse poly-immunoglobulin receptor, cDNA.
KW heavy chain; antigen binding domain; protection protein; mutans;
KW mucosal; environment; gastrointestinal; passive; immunisation;
KW Guy's 13 antibody; prevention; dental caries; Streptococcus;
KW poly; sorbinus; murine; ss.
OS Mus musculus.
FH Key Location/Qualifiers
FT cds 85..2400
FT /*tag= a
PN WO9621012-A1.
PD 11-JUL-1996.
PF 27-DEC-1995; U16889.
PR 30-DEC-1994; US-367395.
PR 04-MAY-1995; US-434000.
PA (PLAN-) PLANT BIOTECHNOLOGY INC.
PA (UNNE-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.
PA (PLAN-) PLANET BIOTECHNOLOGY INC.
PI Hiatt AC, Lehner T, Ma JKC;
DR WPI: 96-333987/33.
DR P-PSDB: W03180.
PT Immunoglobulin and protection protein complex and its prodn. in
PT plants - usefui for passive immunisation against mucosal antigens,
PT esp. against S. mutans and S. sorbinus to prevent dental caries
PS Disclosure: Pages 117-121; 152pp; English.
CC The present sequence encodes the mouse poly-immunoglobulin (Ig)
CC receptor, a portion of which corresp. to residues 1-627, pref.
CC 1-606 or esp. residues 13-45, 1-120, 110-230, 210-340, 320-450,
CC 440-550, 550-606 or 550-627 comprises a protection protein (PP).
CC The Ig of the invention comprises a PP as above in association with
CC an Ig derived heavy chain, having at least a portion of an antigen
CC (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.
CC gastrointestinal, environments, therefore enhancing its
CC effectiveness in passively immunising animals against mucosal
CC pathogens. The Ag binding domain is specifically derived from the
CC Guy's 13 antibody, and the Ig can be used to prevent dental caries
CC by binding, e.g. Streptococcus mutans serotypes c, e and f, or
CC S. sorbinus serotypes d and g.
SQ Sequence 3095 BP; 861 A; 796 C; 784 G; 654 T;

alignment_scores:
Quality: 198.00 Length: 168
Ratio: 1.980 Caps: 5
Percent Similarity: 59.524 Percent Identity: 31.548

alignment_block:
US-09-135-238B-2 x T31290 ..
Align seg 1/1 to: T31290 from: 1 to: 3095
10 PheLeuProValSerGlyAlaLeuArgIleLeuProGluValLysValG1 26
121 TTTTACGGGGTCTCCACAAAAGCCCATATTTGTCCTCCAGAGGTGAG 170
26 uGlyGluLeuGlyGlySerValThrIleLysCysProLeuProGlu... 41
171 TAGTATAGAAGGCGACTCTGTTTCCATCAGGTGCTACTACCCAGACACT 220
42 .....MetHisValArgIleTyrLeuCysArgGluMetAlaGlySer 55
221 CTGTCAACCGGCACACCGGAAATATCTGTCGCCCAAGAGGCC...AGC 267
56 GlyThrCysGlyThrValValSerThrThrAsnPheIleLysAlaGluT 72
268 GGCATGTGCACACGCTCATCTCTTCAATGGCTACCTCTCCAGGAGTA 317
72 rLysGlyArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeu 89

```

```

318 TTCAGCAGAGCAACCTCATCACTCCAGAGAACACACATTGTGA 367
89 algluValThrGlnLeuThrGluSerAspSerGlyValThrAlaCysGly 105
      ::::: ||||| ::::: ||||| ::::: |||||
368 TTAACATGAGCAGCTCACCCAGGACGACACTGGAGCTACAAGTGTGGC 417
      ||| ::::: ||||| ::::: |||||
106 AlaGlyMetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnVa 122
      ||| ::::: ||||| ::::: |||||
418 CTGGGT...ACAGTAACCGAGGCTGCTTCGATGTCAGCTGGAGGT 464
      ||| ::::: ||||| ::::: |||||
122 lHisSerGluTyrGluProSerTyrGluGluGlnProMetProGluThrP 139
      | ::::: |||||
465 C.....AGCCAGGTCTCTGAGTGC 484
139 rolystrpPheHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSer 155
      ||| ::::: |||||
485 CGAGTGACACCCAGTC.....TACACAAAG 510
156 SerSerLysPheValThrArgValThrThrProAlaGlnArgGlyLysVa 172
      :: ::::: |||||
511 GACATAGCAGAAATGTGACCATTTGAATGCCCTTTTCAAAGGGAGAAATGT 560
172 lPro 173
      ||||
561 TCCC 564

seq_name: N_Geneseq_36.V15422
seq_documentation_block:
ID V15422 standard; DNA; 29392 BP.
AC V15422;
DT 11-JUN-1998 (first entry)
DE Mouse poly Ig receptor protein gene.
KW Mouse; poly Ig receptor protein; pIgR protein; pIg; deficiency;
KW knockout mouse; disease model; ds.
OS Mus sp.
PN J10057066-A.
PD 03-MAR-1998.
PF 19-AUG-1996; 217154.
PR 19-AUG-1996; JP-217154.
PA (HONS) TAKUJI HONSHA KK.
DR WPI; 98-254323/23.
PT Mouse pIg receptor protein gene - used for preparing gene knockout
PT mice, useful for study of human poly Ig receptor protein deficiency
PS Claim 1; Page 4-14; 18pp; Japanese.
CC The present sequence represents the mouse poly Ig receptor protein
CC gene, which has a 29392 bp sequence. The new gene can be used to
CC produce a gene knockout mouse, useful as a disease model of human
CC poly Ig receptor protein deficiency.
SQ Sequence 29392 BP; 8318 A; 6747 C; 6514 G; 7813 T;

alignment_scores:
      Quality: 188.00      Length: 97
      Ratio: 2.648      Gaps: 3
Percent Similarity: 73.196      Percent Identity: 40.206

alignment_block:
US-09-135-238B-2 x V15422 ..
Align seg 1/1 to: V15422 from: 1 to: 29392

30 GlyGlySerValThrIleLysCysProLeuProGlu.....Me 42
18245 GCGGACTGTGTTCCATACAGTGTGCTACTACCCAGACACCTCTGTCAACCG 18294
42 tHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 59
18295 GCACACCCGGAAATGCTGTGTCGACAAAGAGCC...AGCGCATGTGCA 18341
59 lYthrValValSerThrAsnPheIleLysAlaGluTyrLysGlyArg 75
18342 CACCGCTCATCTCTTCAATGGCTACTCTCCAAAGGAGTATTCAGGCAGA 18391

```

```

76 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 92
      ::::: ||||| ::::: ||||| ::::: |||||
18392 GCCAACCTCATCAACTTCCAGAGAACACACATTGTGATTAACATTGA 18441
      ||||| ::::: ||||| ::::: |||||
92 rGlnLeuThrCluserAspSerGlyValThrAlaCysGlyAlaGlyMeta 109
      ||||| ::::: ||||| ::::: |||||
18442 GCAGCTCACCCAGAGACACACTGGAGCTACAAGTGTGGCTGGGT...A 18486
      ::::: ||||| ::::: |||||
109 snThrAspArgGlyLysThrGlnLysValThrLeuAsnVal 122
      ::::: ||||| ::::: |||||
18489 CCAGTAACCCAGGCTGCTTCGATGTCAGCTGGAGGTC 18529
seq_name: N_Geneseq_36.V20383
seq_documentation_block:
ID V20383 standard; cDNA to mRNA; 1911 BP.
AC V20383;
DT 26-JUN-1998 (first entry)
DE cDNA for human immunity related factor.
KW Lymph node; human; immunity related factor; research; treatment;
KW immune disease; infectious disease; ds.
OS Homo sapiens.
FH Key
FT CDS
      Location/Qualifiers
      98..1696
      /tag= a
      FT sig_peptide
      98..145
      /tag= b
      FT mat_peptide
      146..1693
      /tag= c
      /product= immunity_related_factor
J10072495-A.
PD 17-MAR-1998.
PF 11-JUN-1997; 153218.
PR 13-JUN-1996; JP-152362.
PA (ASAH) ASAH KASEI KOGYO KK.
DR WPI; 98-234766/21.
DR P-PSDB; W50033.
PT Immunity related factor - useful in the treatment of immune related
PT and infectious diseases
PS Claim 15; Pages 18-20; 21pp; Japanese.
CC The present sequence encodes a lymph node derived human immunity
CC related factor, which can be used to research and treat immune and
CC infectious diseases.
SQ Sequence 1911 BP; 490 A; 541 C; 525 G; 355 T;

alignment_scores:
      Quality: 187.50      Length: 575
      Ratio: 0.906      Gaps: 20
Percent Similarity: 36.000      Percent Identity: 18.957

alignment_block:
US-09-135-238B-2 x V20383 ..
Align seg 1/1 to: V20383 from: 1 to: 1911

3 ArgTyrLeuTyrProLeuTyrPheLeuProValSerGlyAlaLeuArgTil 19
170 AGATGGCTGTGGAG...GGCTCTCTCCCTCCAGGACCCATCTCCGGGC 216
19 eLeuProGluValLys..... 24
217 CATGGAAACACTCAGGCTCTCTCCCTCTCTGCGGGGAGGAGCT 266
25 .....ValcLuglyVlu 28
267 CCTTTGCAGCTCCAAATTCATTGAAGGCTCAAGGCTGTGTCTCAGGGAG 316
29 LeuGlyGlySerValThrIleLysCysProLeuPro.....G1 41
317 CTGGAGGAGCTGTACCATTCAGGATTCAGGATTCATCTGTCAA 366

```

```

41  uMethHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrC 58
   :   |||   |||   |||   |||   |||   |||   |||   |||
367 CAGGCACACAGAGAGTACTGTCGCCGTCGGGCCCCCAAGATGGATCT 416
58  ysGlyThrValValSerThrThrAsnPhelLeuLysAlaGluTyrLysGly 74
   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||
417 GCCAGACCATTTGTCCACCAACCAAGTATACCTACCACTCGCTATCGTGAC 466
75  ArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluVa 91
   |||   |||   |||   |||   |||   |||   |||   |||   |||
467 CGTGTGGCCCTCAGAGACTTCCACAGAGAGGCTGTTGTGTGGAGGCT 516
91  lThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyM 108
   :   |||   |||   |||   |||   |||   |||   |||   |||
517 GTCCCAACTGTCCCGGATGACATCGATGCTAGCTCTGCGGCATTGGAA 566
108 eTAsnThrAspArg..... 112
567 GTGAAACAACATGCTGTCTTAAGCATGAATCTGACCATCTCTGCAGGT 616
112 ..... 112
617 CCGCCAGCACCCCTCCACACGCCACTCCAGCTGCTGGGGAGCTCACCAT 666
113 .....GlyL 114
667 GAGATCTTATGGAACAGAGCTCTCCAGTGGCCACAGATGGACCCAGAA 716
114 ysThrGln..... 116
717 CCACCCAGACCTTAGGACAGGGGACAGCATGGACACAGTTGCTTCCACT 766
117 .....LysValThrLeuAsnValHisSerGluTyrGluPro... 128
767 CCAGGAAACCAGACAGACTACAGCTTCAGCTGAGGAAGACCAACCCAGG 816
129 .....serTrpGluGlu..... 132
817 AGCAACAGCGCCAGCAGCTCCAGGACAGGAGGCTGGCGCAGGGTCTG 866
133 .....GlnProMetProGluThrProLysTyrPheHisLeuPro 145
867 TCAAGCACCTGCTCCGATTCAGAGAGTCCA..... 898
146 TyrLeuPheGlnMetProAlaTyrAlaSerSerLysPheValThrAr 162
899 .....CCTTCAAAAGACAGACAGCATGTCCAA 924
162 gValThrThrProAlaGlnArgGlyLysValProProValHisSers 179
925 TACAACAGAAAGTGTGTGGAGGGC..... 949
179 exProThrThrGlnIleThrHisArgProArgValSerArg..... 192
950 ..ACCAGAGCTCGGTGACAAACAGGCTAGAGCCAGCAGGACGAGGAGG 997
193 ..AlaSerSerValAlaGlyAspLysProArg..... 202
998 GAGATGACAACCTACCAAGGCTGATAGGCCAAGGAGGACATAGAGGGGT 1047
202 ..... 202
1048 CAGGATAGCTCTGTATGATGACGCCAAAAGGCTCTAGGAACCATTTGGGCCAC 1097
203 .....ThrPheLeuProSerThrThr 209
1098 CAGCTCTGGTCTCAGAAACTTTGGCTGGGAAATCCTCCCAAGCAACG 1147
210 AlaSerLysIleSerAlaLeuGluGlyLeuLeuLysProGlnThrProSe 226
1148 CCAATTCTTAAGCAACAATCTCAGGGTCTCCATTTGGAGAAACAATCCAGC 1197
226 rTyAsnHisHisThrArg..... 232

```

```

1198 TGCAGGCAATGTGGACCTTGGAACTCGCACTGCGATGTGTGATCTTGG 1247
233 .....LeuHisArgGlnArgAlaLeuAspTyr 241
1248 GAATCCAGCTGCAGATGTGTGGACCAAGCATGGAGCGAGCATCTGGGAA 1297
242 GlySerGlnSerGly..... 246
1298 GGAAGCGCTGAGGGACCTAGATGCTGCCACTGGAGACAGAGGTCCCCA 1347
246 ..... 246
1348 AGCAACACTGAGCCAGACCCCGGAGTAGGACCTGGGGACCCCTGGCA 1397
247 .....ArgGluGlyGlnGlyPheHis 253
1398 AGGAGTCTCCGTGAGGGTACTTTCCAGAAGATGAAAGCAGCTCTCGG 1447
254 lIeLeuIlePro.....ThrIleLeuGlyLeuPheLeuLeuLeu 268
1448 ACCCTGGCTCTGCTCTACCATGCTGGCCCTGTTATGCTTATGGCTCT 1497
268 uGlyLeuValValLysArgAlaValGluArgLysAlaLeuSerArg 285
1498 GGTCTATTG.....CAAAGGAGCTCTGGAGAGGA 1529
285 rgAlaArgArgLeuAlaValArgMetArgAlaLeuGluSerSerGln... 300
1530 GGACCTCTCAGGAGGACAAAGGTCACCTTAATTCAGATGACACATTTT 1579
301 .....ArgProArgGlySerProArgProArgSerGlnAsnAsnI 314
1580 CTGGAAGTGAACCCCAAGCAGACCACTGCCCTGTGGAAGAAGAT 1629
314 eTyr.....SerAlaCysProArgArgAlaArgGlyAlaAspAlaAg 329
1630 GCTCCAGGATGACTCTCTCT.....GCTGGGGCCAGCC 1664
329 lYThrGlyGluAlaProValProGlyPro.....Gly 339
1665 TGACTGCCCCAGAGAAATCCAGGACCTTGAGGACAGAGATGAACT 1714
340 AlaProLeuPro.....ProAl 345
1715 GCTCAGTTTACCATGGAGAGGACCAAGATCAAAGCCTTCAGAGCCCA 1764
345 aProLeuGlnValSerGluSerPro 353
1765 GCCTCTTCCATCATCTCTCTCTCA..... 1789
seq_name: N_Geneseq_36:T31291
seq_documentation block:
ID T31291 standard; cDNA; 3269 BP.
AC T31291;
DT 24-FEB-1997 (first entry)
DE Rat poly-immunoglobulin receptor, cDNA.
KW Rat; immunoglobulin; receptor; protection protein; mutants;
KW heavy chain; antigen binding domain; protection; pathogen;
KW mucosal; environment; gastrointestinal; passive; immunisation;
KW Guy's 13 antibody; prevention; dental caries; Streptococcus;
KW poly; sorbinus; ss.
OS Rattus rattus.
Location/Qualifiers
FH key 74..2383
FT cds /*tag= a
FN W09621012-A1.
PD 11-JUL-1996.
PF 27-DEC-1995; U16889.
PR 30-DEC-1994; US-367395.
PR 04-MAY-1995; US-434000.
PR (PLAN-) PLANT BIOTECHNOLOGY INC.

```







```

74 yArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluV 91
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
255 CAGGGCTAACCTCAACCACTTCCCGGAGAACGGCAGCATTCGTGGTGAACA 304
91 aThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGly 107
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
305 TTGCCACGCTGAGCCAGAGTACCTCCGGCGCTACAAGTGTGCCCTGGGC 354
108 MetAsnThr..... 110
355 ATCAATAGCCGAGGCGCTGTCTTTGATGTACGCTGGAGGTACGCCAGG 404
111 .....AspArgGlyLysT 115
405 TCCTGGGCTCTTAATGACACTAAGTCTACACAGTGGACCTGGCGAGAA 454
115 hrGlnLysValThrLeuAsnValHisSerGluTyrGluProSerTrpGlu 131
|||:|||||:|||||:|||||:|||||:|||||:|||||:
455 CG.....GTGACCATCAACTGCCCTTTCAAGACTGAGATGCTCAAAAG 498
132 GluGln.PrometProGlu.....ThrProLysTrpPheHisL 144
|||||:|||||:|||||:|||||:|||||:|||||:
499 AGAAGTCTTGTACACACATAGGCGTGTACCCCTGTCTGTCTATCGA 548
144 euProTyrLeuPheGlnMetProLysTyrAlaSerSerLysPheVal 160
|||:|||||:|||||:|||||:|||||:|||||:|||||:
549 CTCAGTGTGTATGTGAATCCCACTATACAGGAAGAATACGCTTGATA 598
161 ThrArgVal..... 163
599 TTCAGGGTACTGGCCAGCTACTGTTCAGCGTGTGCATCAACCACTCAGG 648
164 .....ThrThrProAlaGlnArgGlyLysValProProV 175
|||:|||||:|||||:|||||:|||||:|||||:|||||:
649 CTCAGCGATGCTGGCAGTATCTCTGCCAGCGTGGGATGATTCCTCAATAG 698
175 aHisHisSerSerProThrThrGlnLeuThrHisArgProArgValSer 191
|||:|||||:|||||:|||||:|||||:|||||:|||||:
699 TAATAGAGAAATGCTGACC..... 718
192 ArgAlaSerSerValAlaGlyAspLysProArgThrPheLeuProSerTh 208
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
719 .....TCCAAGTGTAAAGCCGAGCGCGAGCTGTTTATGAAGACCTC 762
208 rThrAlaSerLysLleSerAlaLeuGluGlyLeuLysProGlnThr. 224
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
763 AGGGCTAGTACCTTCCACTGTGCCCTGGCCCTGAGTGGCAACGT 812
225 ..ProSerTyrAsnHisHisThrArgLeuHisArg..... 235
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
813 GGCCAAATTTCTGCCGACAGACGAGTGGGAAACACTGTGACGTGGTCG 862
236 .....GlnArgAlaLeuAspTyrGlySerGlnSe 245
863 TCACACCTTAGGGAAGAGGCCCGCCCTTGAGGCGAGGATCCTGCTC 912
245 r..... 246
913 AACCCCCAGGACAGGCTGCTTTCAGTGTGTGATCATCAGGCGCTGAG 962
246 lyArgGluGlyGlnGlyPheHisLleLeuIleProThrIleLeuGlyLeu 262
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
963 GAAGGAGGATGCGAGGCGCTACCTGTGTGGAGCCCAT. 1000
263 PheLeuLeuAlaLeuLeuGlyLeuValLysArgAlaValGluArgAr 279
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1001 .....CGGATGCTCAGCTGCAG 1017
279 glyAlaLeuSerArgAlaArgArgLeuAlaValArgMetArgAlaL 296
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1018 GAAGGCT.....CGCCTATCCAGGC 1037

```

```

296 euGluSerSerGlnArgProArgGlySerProArgProArgSer..... 310
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1038 CTGGCAACTCTTCGTCAATGAGGAGTCCACCATTCGCCCGAGCCCACTG 1087
311 .....GlnAsnAsnIleTyr.....SerAlaCysProArgAr 321
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1088 TGGTGAAGGGGTGGCAGGAGCTCTGTGGCGCTGCTGCCCTACAAC 1137
321 galaArgGlyAlaAspAlaAlaGlyThrGlyGluAlaProValProGlyP 338
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1138 CGTAAGGAAGCAAAAGCATCAAGTACTGGT.....GTCTCTGGGA 1178
338 roGlyAlaProLeuProProAlaPro..... 346
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1179 AGGGCCCCAGATGGCGCTGCCCTCTGCTGTGGACAGCGGGGTGGG 1228
347 .....LeuGlnValSerGluSerProTrp.....LeuH1 356
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1229 TTAAGGCCACTACGAGGGCGCGCTCTCCCTGTGGAGGAGCCAGGCAAC 1278
356 sAlaProSerLeuLysThrSer 363
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1279 GGCACCTTCACTGTCTCATCTCA 1300
seq_name: N_Geneseq_36:X07407
seq_documentation_block:
ID X07407 standard; cDNA to mRNA; 1839 BP.
AC X07407;
DT 08-JUN-1999 (first entry)
DE Human secretory immunoglobulin A component gene.
KW immunoglobulin A; secretory; component; iGa; human; treatment;
KW prevention; infection; HIV; AIDS; cold; flu; virus; gene;
KW human immunodeficiency virus; respiratory syncytial virus; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 1..1839
FT /tag= a
FT /product= secretory component
PN W09857993-A1.
PD 23-DEC-1998.
PF 10-JUN-1998; U11975.
PR 19-JUN-1997; US-050969.
PA (REGC ) UNIV CALIFORNIA.
PI Chintalacharuvu KR, Morrison SL;
DR WPI: 99-080950/07.
DR P-FSDB; W95601.
PT Producing secretory immunoglobulin in single cells - useful to
PT produce commercial quantities of secretory immunoglobulin to prevent
PT or treat infections
PS Disclosure; Page 22; 39pp; English.
CC The sequence is that of the coding region for the secretory
CC component of human secretory immunoglobulin A (sIgA).
CC The sequence can be used as part of a method for the
CC production of sig molecules. This method is useful for
CC producing commercial quantities of sig (especially sIgA) to treat
CC or prevent infections. In particular, sIgA produced by the method
CC can be used to prevent or treat infections in mammals, birds or
CC fish; especially systemic infections or infections at a mucosal
CC surface. It is especially useful to prevent or treat infection
CC with human immunodeficiency virus (HIV), respiratory syncytial
CC virus, flu virus or cold virus. The method allows production of
CC commercial quantities of sig molecules for therapeutic use, not
CC previously possible; production using non-plant cells and a
CC single cell type is more efficient than a previous multi-step
CC process of fusing recombinant plant cells, and avoids alterations
CC of the sig by plant cells. sIgA molecules are more stable
CC and resistant to proteolysis than previously used IgA molecules,
CC and can be administered to prevent as well as to treat infections,
CC unlike e.g. IgG and IgM molecules.
CC Sequence 1839 BP; 423 A; 504 C; 544 G; 368 T;
SQ

```

```

alignment_scores:
  Quality: 182.00      Length: 115
  Ratio: 2.247        Gaps: 5
  Percent Similarity: 70.435  Percent Identity: 37.391

alignment_block:
  US-09-135-238B-2 x X07407  ..
  Align seg 1/1 to: X07407 from: 1 to: 1839

12  ProValSerGlyAlaLeuArgIleLeuProGluValLysValGluGly 28
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61  CCATATTGGT.....CCCGAG...GAGGTGAATAGTGT 92
28  uLeuGlySerValThrIleLysCysProLeuPro.....G 41
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
93  GGAAGTAACCTACGTGTCATCAGTGTCTACCCACCCACCTCTGTCA 142
41  LuMetHisValArgIleLeuLysCysArgGluMetAlaGlySerGlyThr 57
    :: |||:|||||: ||| |||:|||||: ||| |||:
143  ACCGGCACACCCGGAAGTACTGTCGCCGCGAGGAGCTAGAGTGGC... 189
58  CysGlyThrValValSerThrAsnPhelLysAlaGluTyrLysGI 74
    ||| |||:|||||:|||||:|||||:|||||:|||||:|||||:
190  TGCATACCCCTACCTCTCGGAGGGCTACGTCTCCAGCAATATGCGG 239
74  YArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGlu 91
    |||:|||||: ||| |||:|||||: ||| |||:|||||:
240  CAGGGCTAACCTCACCACACTTCCCGGAGAACGCCACATTTGTGTGAACA 289
91  alThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGly 107
    :: |||:|||||: ||| |||:|||||: ||| |||:|||||:
290  TTGCCACACTGACGAGGATGACCCGGCGCTACAAAGTGTGGCTGGGC 339
108  MetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnVal 122
    |||:|||||: ||| |||:|||||: ||| |||:|||||:
340  ATCAATAGC...CGAGGCTGTCTTGTGATGTGTCAGCTGGAGGTC 381

seq_name: N.Geneseq_36:T30856

seq_documentation_block:
ID  T30856;
AC  T30856 standard; cDNA; 2031 BP.
DE  13-SEP-1996 (first entry)
DE  Secretory component DNA fragment 1.
KW  Secretory component; SC DNA fragment 1; crystallisation;
KW  Polymeric immunoglobulin receptor; pigR; CHO; ss.
OS  Chimeric Homo sapiens;
OS  Chimeric synthetic.
FH  Key
FH  Location/Qualifiers
FT  misc_feature
FT  1..33
FT  /*tag= a
FT  /note= "bases 1-33 derived from PCR with 5'
FT  primer 1"
FT  111..2015
FT  /*tag= b
FT  /product= transcript from SC DNA fragment 1
FT  complement (1997..2031)
FT  /*tag= c
FT  /note= "from PCR with 3' primer"
FT  WO9618734-A1.
PN  20-JUN-1996.
PD  06-DEC-1995; E04797.
PR  16-DEC-1994; EP-120019.
PA  (CIBA ) CIBA GEIGY AG.
PI  Asselbergs FAM, Gruetter M, Leist C, Van Oostrum J;
DR  WPI; 96-300651/30.
PT  prodn. of recombinant secretory protein capable of crystallisation -
PT  pref. human poly-IgR. for use in screening and binding studies
PS  Claim 5; Page 36-38; 50pp; English.
CC  SC DNA fragment 1 (T30856) was obtd. by PCR amplification of human
CC  polymeric immunoglobulin receptor (pigR) cDNA in vector pCB6 using a
CC  5' primer (T30859) hybridising to the cytomegalovirus promoter in

```

```

CC  pCB6 and a 3' primer (T30861) that generates a stop codon at
CC  position 1906 just upstream of the hydrophobic transmembrane segment
CC  of the pigR and also creates a downstream XbaI site. CHO SSF 3
CC  cells transfected with a vector carrying the amplified DNA can be
CC  used for prodn. of recombinant secretory component (SC). This can
CC  be crystallised for receptor structure studies, used to stabilise
CC  immunoglobulins, or used to screen (ant)agonists capable of
CC  modulating mucosal immune responses.
SQ  Sequence 2031 BP; 477 A; 541 C; 605 G; 408 T;

alignment_scores:
  Quality: 182.00      Length: 115
  Ratio: 2.247        Gaps: 5
  Percent Similarity: 70.435  Percent Identity: 37.391

alignment_block:
  US-09-135-238B-2 x T30856  ..
  Align seg 1/1 to: T30856 from: 1 to: 2031

12  ProValSerGlyAlaLeuArgIleLeuProGluValLysValGluGly 28
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
171  CCATATTGGT.....CCCGAG...GAGGTGAATAGTGT 202
28  uLeuGlySerValThrIleLysCysProLeuPro.....G 41
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
203  GGAAGTAACCTACGTGTCATCAGTGTCTACCCACCCACCTCTGTCA 252
41  LuMetHisValArgIleLeuLysCysArgGluMetAlaGlySerGlyThr 57
    :: |||:|||||: ||| |||:|||||: ||| |||:|||||:
253  ACCGGCACACCCGGAAGTACTGTCGCCGCGAGGAGCTAGAGTGGC... 299
58  CysGlyThrValValSerThrAsnPhelLysAlaGluTyrLysGI 74
    ||| |||:|||||:|||||:|||||:|||||:|||||:|||||:
300  TGCATACCCCTACCTCTCGGAGGGCTACGTCTCCAGCAATATGCGG 349
74  YArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGlu 91
    |||:|||||: ||| |||:|||||: ||| |||:|||||:
350  CAGGGCTAACCTCACCACACTTCCCGGAGAACGCCACATTCGTGTGAACA 399
91  alThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGly 107
    :: |||:|||||: ||| |||:|||||: ||| |||:|||||:
400  TTGCCACACTGACGAGGATGACCCGGCGCTACAAAGTGTGGCTGGGC 449
108  MetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnVal 122
    |||:|||||: ||| |||:|||||: ||| |||:|||||:
450  ATCAATAGC...CGAGGCTGTCTTGTGATGTGTCAGCTGGAGGTC 491

seq_name: N.Geneseq_36:T31288

seq_documentation_block:
ID  T31288 standard; cDNA; 2919 BP.
AC  T31288;
DE  24-FEB-1997 (first entry)
DE  Human poly-immunoglobulin receptor, cDNA.
KW  Human; immunoglobulin; receptor; protection protein; mutants;
KW  heavy chain; antigen binding domain; protection; pathogen;
KW  mucosal; environment; gastrointestinal; passive; immunisation;
KW  Guy's 13 antibody; prevention; dental caries; Streptococcus;
KW  poly; sorbinus; ss.
OS  Homo sapiens.
FH  Key
FH  Location/Qualifiers
FT  cds
FT  235..2475
FT  /*tag= a
FT  WO9621012-A1.
PN  11-JUL-1996.
PD  27-DEC-1995; U15889.
PR  30-DEC-1994; US-367395.
PR  04-MAR-1995; US-434000.
PA  (PLAN-) PLANT BIOTECHNOLOGY INC.
PA  (UNNE-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.
PA  (PLAN-) PLANET BIOTECHNOLOGY INC.

```



119 rIeuAsnValHisSerGlu 125  
 :||||:|||| :|||  
 523 CTGGAGGTGAGCAAGAT 541

seq\_name: N\_Geneseq\_36:T96033

seq\_documentation\_block:

ID T96033 standard; CDNA; 2084 BP.

AC T96033;

DT 21-MAY-1998 (first entry)

DE Rat kidney injury related molecule (KIM) cDNA clone 1-7.

KW Kidney injury related molecule; KIM; rat; renal disease; injury;

KW Nephritis; tissue regeneration; therapy; ss.

OS Rattus sp.

FH Key Location/Qualifiers

FT CDS 145..1068

FT /\*tag= a

PN W09744460-A1.

PD 27-NOV-1997.

PF 23-MAY-1997; 009303.

PR 23-AUG-1996; US-023442.

PR 24-MAY-1996; US-018228.

PA (BIOJ) BIOGEN INC.

PI Bonventre JV, Cate RL, Hession CA, Ichimura T, Sanicola-Nadel M,

PI Wei H;

PI WPI: 98-018514/02.

DR P-PSDB; W38334.

PT DNA encoding kidney injury related molecule - which is upregulated in injured or regenerating tissue, useful to promote growth of new tissue and survival of damaged tissue

PS Claim 1; Page 34-37; 68pp; English.

CC cDNA clone 1-7, deposited as ATCC 98060, codes for a rat protein (see W38334), designated kidney injury related molecule (KIM), that is up-regulated in injured or regenerating tissue. Representational

CC difference analysis was used to examine cDNA libraries prepared from

CC ischaemic and normal rat adult kidneys. 3 Cycles of hybridisation

CC and selective amplification gave 3 fragments present only in the

CC injured kidney library. Screening of the injured kidney library

CC with these fragments gave clone 1-7, as well as clone 3-2 (see

CC T96032), which is a splice variant of 1-7 encoding an identical KIM

CC protein (nucleotides 136-605 of 3-2 represent an insertion), and

CC clone 4-7 (see T96034) encoding a 307-amino acid KIM protein (see

CC W38335). A labelled probe from clone 1-7 was used to identify a

CC human KIM sequence (see T96035). KIM nucleic acids can be used in

CC the recombinant production of KIM polypeptides in prokaryotic or

CC eukaryotic host cells. KIM, or an agonist, can be used to treat renal

CC disease and to promote the growth of new tissue or the survival of

CC damaged tissue, generally in conditions where the binding of

CC specific ligand to KIM stimulates cell growth, maintains cellular

CC differentiation or reduces apoptosis, e.g. in cases of renal

CC failure, nephritis, kidney transplants, toxic or hypoxic injury.

CC Damage/regeneration of renal cells can be determined by measuring

CC KIM polypeptide or nucleic acid, e.g. with an antisense probe,

CC particularly to diagnose or monitor the progress of disease or

CC therapy.

SQ Sequence 2084 BP; 604 A; 451 C; 470 G; 559 T;

alignment\_scores:

Quality: 171.00 Length: 403

Ratio: 0.924 Gaps: 22

Percent Similarity: 45.906 Percent Identity: 24.318

alignment\_block:

US-09-135-238B-2 x T96033 ..

Align seq 1/1 to: T96033 from: 1 to: 2084

13 ValSerGlyAlaLeuArgIleLeuPro.....GluVa 23  
 :||||| :||| :|||

166 ATTCAGGCTCTGCTGCTCTTCCAGGCTCTGTAGATTCATTATGAAGT 215

23 LysValGluGlyLeuGlySerValThrIleLysCysProLeuP 40

1 :||||| :||| :||| :||||| :|||  
 216 A...GTGAAGGGGTGGTGGGTCAACCTGTCACAATTCATGCTACTTACT 262  
 40 roGluMetHisValArgIleTyrLeuCys.....ArgGluMetAlaGly 54  
 :||| :||| :||| :||| :|||  
 263 CAAACAGTGGAGGAATCACACAGCATGTTGGGGCGGGGCAATGCCCA 312  
 55 SerGlyThrCysGlyThrValValSerThrThrAsnPhelIleLysAlaG 71  
 :||||| :||||| :||||| :||||| :|||||  
 313 TATTCTAGTTGTCAAATATATCTTATTGGACCAATGGATACCAAGTCAC 362  
 71 uTyrLys.....GlyArgValThrLeuLysGlnTyrProArgLysAsnL 86  
 :||||| :||||| :||||| :||||| :|||||  
 363 CTATCGGAGCAGCGGTGATACACATAAAGGGCGGTATTTCAGAAGGAG 412  
 86 euPheLeuValGluValThrGlnLeuThrGluSerAspSerGlyValTyr 102  
 :||| :||| :||| :||| :|||  
 413 ACGTATCTCTGCACANTAGAGAACTCTGTGTATAGTAGTAGTGGTCTGAT 462  
 103 AlaCysGlyAla.....GlyMetAsnThrAspArgGlyLysThrG 116  
 :||| :||| :||| :||| :|||  
 463 TGTGTCGAGTGGAGATCTCTGGATGTTCAACGAT.....CA 500  
 116 nLysValThrLeuAsnValHisSerGluTyrGluProSerTrpGluGluG 133  
 :||||| :||||| :||||| :||||| :|||||  
 501 GAAATGACCTTTTCATG.....GAAGTTAAACCAAGAAATTCACCAA 544  
 133 lnProMetProGluThrProLysTrpPheHisLeuProTyrLeuPheGln 149  
 :||| :||| :||| :||| :|||  
 545 GTCT....CAACAAAGACCC..... 561  
 150 MetProAlaTyrAlaSerSerSerLysPheValThrArgValThrThrPr 166  
 :||||| :||||| :||||| :||||| :|||||  
 562 .....ACAACACAGAGCCCAACCAACCAAGGCCCAACTAT 599  
 166 oAlaGlnArgGlyLysValProPro.....ValHisHisSerS 179  
 :||| :||| :||| :||| :|||  
 600 TTCACAGATCCACATGTCACCAACATCAACACAGAGTCCACCTCTA 649  
 179 exProThr.....ThrGlnIleThrHisArgProArgValSerArg 192  
 :||||| :||||| :||||| :||||| :|||||  
 650 CTCACACACCAAGAAACACAG.....ACTCACAACCAAGAAATCACT... 693  
 193 AlaSerSerValAlaGlyAspLysProArgThrPheLeuProSerThrTh 209  
 :||||| :||||| :||||| :||||| :|||||  
 694 .....ACATTTATGCCCATGAC 713  
 209 rAlaSerLysIleSerAlaLeuGluGlyLeuLysProGlnThrProS 226  
 :||||| :||||| :||||| :||||| :|||||  
 714 AACTGCTGAGGTGACA.....GAAACTCCAT 739  
 226 erTyr..... 227  
 :||||| :||||| :||||| :||||| :|||||  
 740 CATATATCTCTGCAGACTGGAATGCCACTGTCATCTCCTCAGAGGAGGCC 789  
 228 .....AsnHisHisThrArgLeuHisArgGlnAlaLeuAspTyrG 242  
 :||||| :||||| :||||| :||||| :|||||  
 790 TGAATAATCACTGTAAGAATCCCTTTGAGGAAG..... 825  
 242 ySerGlnSerGlyArgGluGlyGlnGlyPheHisIleLeuIleProThrI 259  
 :||| :||| :||| :||| :|||  
 826 ....CGCAGAGAAACCCGACTAAGGGCTCTATGTTGGCATGTCCGTTG 871  
 259 leLeuGlyPheLeuLeuAlaLeuLeuGlyLeuValValLysArgAla 275  
 :||||| :||||| :||||| :||||| :|||||  
 872 CAGCCCTGCTCTCTCTGCTTGGCAGCACCGGTGTTGTCACAGGATAC 921  
 276 ValGluArgArgLys.....AlaLeuSerArgArgAlaArgArgLe 289  
 :||| :||| :||| :||| :|||  
 922 ATCATTATAAGAAAGAGATGGGCTCTCTGAGCTTTGCTGCTCCATGT 971  
 289 uAlaValArgMetArgAlaLeuGluSerGlnArgProArgGlySerP 306  
 :||| :||| :||| :||| :|||

```

972 CTCT...AAGAGTAGAGCTTTGCAGAACGCA.....GCGATTG 1006
306 roArgProArgSerGlnAsnAsnIleTyrSerAlaCysProArgAla 322
   ::::::::::::::::::::::::::::|
1007 TGCATCCCGAGCTGAAGCAACATCTAC...ATTATTGAAGATAGATCT 1053
   ::::::::::::::::::::|
323 ArgGlyAlaAsp..... 326
   |||
1054 CGAGGTGCAGATAGTCCAGAGGCTTCTGTGGGGCTTCTGCTGGG 1103
327 .....AlaAlaGlyThrClyGluAlaProValProGlyP 338
1104 ATTACAGATCGTACTGATTTCACAGAGTAAATACCAATCCAGCTC 1153
338 roGly.....AlaProLeuProProAlaProLeuGlnValSer 350
   |||
1154 CTGGGAGATTTGTGTTTGTTCCTCCAGCTGCAGTGGAGAGGTAAAC 1203
351 GluSerPro 353
   |||
1204 CTCTACCT 1212

```

seq\_name: N\_Geneseq\_36:T96032

seq\_documentation\_block:

ID T96032 standard; cDNA; 2566 BP.

AC T96032:

DT 21-MAY-1998 (first entry)

DE Rat kidney injury related molecule (KIM) cDNA clone 3-2.

KW Kidney injury related molecule; KIM; rat; renal disease; injury;

KW nephritis; tissue regeneration; therapy; ss.

OS Rattus sp.

FH Key

FT CDS Location/Qualifiers

FT 615..1538

FT /\*tag= a

PN WO974460-A1.

PD 27-NOV-1997.

PF 23-MAY-1997; U093303.

PR 23-AUG-1996; US-023442.

PR 24-MAY-1996; US-018228.

PA (BIOJ) BIOGEN INC.

PI Bonventre JV, Cate RL, Hession CA, Ichimura T, Sanicola-Nadel M,

PI Wei H;

PI WPI; 98-018514/02.

DR P-PSDB; W38334.

PT DNA encoding kidney injury related molecule - which is upregulated

PT in injured or regenerating tissue, useful to promote growth of new

PT tissue and survival of damaged tissue

PS Claim 1; Page 32-34; 68pp; English.

CC cDNA clone 3-2, deposited as ATCC 98061, codes for a rat protein

CC (see W38334), designated kidney injury related molecule (KIM), that

CC is upregulated in injured or regenerating tissue. Representational

CC difference analysis was used to examine cDNA libraries prepared from

CC ischaemic and normal rat adult kidneys. 3 Cycles of hybridisation

CC and selective amplification gave 3 fragments present only in the

CC injured kidney library. Screening of the injured kidney library

CC with these fragments gave clone 3-2, as well as clone 1-7 (see

CC T96033), which is a splice variant of 3-2 encoding an identical KIM

CC polypeptide (nucleotides 136-605 of 3-2 represent an insertion),

CC and clone 4-7 (see T96034) encoding a 307-amino acid KIM protein

CC (see W38335). Clone 1-7 was used to identify a human KIM sequence

CC (see T96035). KIM nucleic acids can be used in the recombinant

CC production of KIM polypeptides in prokaryotic or eukaryotic host

CC cells. KIM, or an agonist, can be used to treat renal disease and to

CC promote the growth of new tissue or the survival of damaged tissue,

CC generally in conditions where the binding of specific ligand to KIM

CC stimulates cell growth, maintains cellular differentiation or

CC reduces apoptosis, e.g. in cases of renal failure, nephritis,

CC kidney transplants, toxic or hypoxic injury. Damage/regeneration

CC of renal cells can be determined by measuring KIM polypeptide or

CC nucleic acid, e.g. with an antisense probe, particularly to

CC diagnose or monitor the progress of disease or therapy.

SQ Sequence 2566 BP; 726 A; 546 C; 591 G; 703 T;

```

alignment_scores:
  Quality: 171.00      Length: 403
  Ratio: 0.924        Gaps: 22
  Percent Similarity: 45.906  Percent Identity: 24.318
alignment_block:
US-09-135-238B-2 x T96032

```

Align seg 1/1 to: T96032 from: 1 to: 2566

```

13 ValSerGlyAlaLeuArgIleLeuPro.....GluVa 23
   ::::::::::|
636 ATTTCAGGCGCTCTGCTCTCTCCAGGCTCTAGATTCTTATGAAGT 685
   ::::::::::|
23 llysValGluGlyLeuGlyGlySerValThrIleLysCysProLeuP 40
   |||
686 A...GTGAGGGGGTGGGTGACCTGTCACAAATCCATGCTACTACT 732
   |||
40 roGluMetHisValArgIleTyrLeuCys.....ArgGluMetAlaGly 54
   |||
733 CAACACGTGGAGGAATACAAACGACATGTTGGGCGCGGGGCAATGCCA 782
   |||
55 SerGlyThrCysGlyThrValValSerThrThrAsnPheIleLysAlaG 71
   ::::::::::|
783 TATTCTAGTTGTCAAAATATATTATTGGACCAATGGATACCAAGTCAC 832
   |||
71 uTyrLys.....GlyArgValThrLeuLysGlnTyrProArgLysAsnL 86
   |||
833 CTATCGGAGCAGCGTCGATACAAATAAAGGCGGTATTTTCAGAAGGAG 882
   |||
86 euPheLeuValGluValThrGlnLeuThrGluSerAspSerGlyValTyr 102
   ::::::::::|
883 ACGTATCTCTTGACAAATAGAGAACTCTGTTGATAGTAGTAGTGTCTGTAT 932
   |||
103 AlaCysGlyAla.....GlyMetAsnThrAspArgGlyLysThrG 116
   ::::::::::|
933 TGTTCCCGAGTGGAGATTCCTGGATGGTTCAACGAT.....CA 970
   |||
116 nLysValThrLeuAsnValHisSerGlyTyrGluProSerTrpGluGlu 133
   |||
971 GAAATGACCTTTTCATTG.....GAAGTTAAACACCAAGAAATCCCAAA 1014
   |||
133 InProMetProGluThrProLysTrpPheHisLeuProTyrLeuPheGln 149
   ::::::::::|
1015 GTCCT...CCACACAGACCC..... 1031
   |||
150 MetProAlaTyrAlaSerSerSerLysPheValThrArgValThrPr 166
   ::::::::::|
1032 .....ACAACCTACAGAGACCCCAACCAACCAAGGCCCAACTAT 1069
   |||
166 oAlaGlnArgGlyLysValProPro.....ValHisHisSerS 179
   ::::::::::|
1070 TTCACACAGATCCACACATGTACCAACATCAACAGAGTCTCCACCTCTA 1119
   |||
179 erProThr.....ThrGlnIleThrHisArgProArgValSerArg 192
   ::::::::::|
1120 CTCCAACACCAACCAACCAACAG...ACTCAACAAACCAAGAAATCACT... 1163
   |||
193 AlaSerSerValAlaGlyAspLysProArgThrPheLeuProSerThrTh 209
   |||
1164 .....ACATTTTATGCCCATGAGAC 1183
   |||
209 rAlaSerLysIleSerAlaLeuGluGlyLeuLysProGlnThrProS 226
   ::::::::::|
1184 AACTGCTGAGGTGACA.....GAAACTCCAT 1209
   |||
226 erTyr..... 227
   |||
1210 CATATACTCTTCAGACACTGGGAATGGCACTGTGACATCTCCTCAGAGGAGGCC 1259
   |||
228 .....AsnHisHisThrArgLeuHisArgGlnArgAlaLeuAspTyrG 242

```



PS Claim 10; Pages 99-102; 152pp; English.  
CC The present sequence encodes the rabbit poly-immunoglobulin (Ig)  
CC receptor, a portion of which corresp. to residues 1-627, pref.  
CC 1-806, or esp. residues 21-43, 1-118, 119-223, 224-332, 333-441,  
CC 442-552, 553-606 or 553-627 comprises a protection protein (PP).  
CC The Ig of the invention comprises a PP as above in association with  
CC an Ig derived heavy chain, having at least a portion of an antigen  
CC (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.  
CC gastrointestinal, environments, therefore enhancing its  
CC effectiveness in passively immunising animals against mucosal  
CC pathogens. The Ag binding domain is specifically derived from the  
CC Guy's 13 antibody, and the Ig can be used to prevent dental caries  
CC by binding, e.g. Streptococcus mutans serotypes c, e and f, or  
CC S. sorbinus serotypes d and g.  
SQ Sequence 3517 BP; 795 A; 1027 C; 1038 G; 657 T;

alignment\_scores:  
Quality: 166.50 Length: 103  
Ratio: 2.602 Gaps: 4  
Percent Similarity: 62.136 Percent Identity: 42.718  
alignment\_block:  
US-09-135-238B-2 x T31287 ..  
Align seg 1/1 to: T31287 from: 1 to: 3517  
30 GlyGlySerValThrIleLysCysProLeuPro.....GluMe 42  
||| |||||:|||| ||| |||  
238 GGCACCTCGGTGCCATCATCATCTACTACCAACACCTCCGTCACCCG 287  
42 tHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 59  
||| ||| :||| ||||| ||||| |||  
288 GCACAGCCGGGAAGTCTGTCGCCGGAA...GAGGAGAGCGCGCGTCG 334  
59 lyThrValIserThrThrAsnPheIleLysAlaGluTyrLysGlyArg 75  
|||:||||| |||:||||| :||| |||||:|||||  
335 TGACGCTTGCCCTCG...ACCGCTACAGTCCACAGGAATACTCCGGGAGA 381  
76 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 92  
||| :||||| |||: |||: ||| |||  
382 GGCAAGCTCACCGACTTCCTGATAAAGGGGAGTTTGTGTGACTGTGA 431  
92 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyWeta 109  
|||||:|||||:|||||:||||| ||| |||||:|||||:|||||  
432 CCAACTACCCAGACACGACTCAGGAGCTACAAGTGTGGCTGGGAGTCA 481  
109 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisserGlu 125  
|| ||||| |||:||||| ||| :|||: ||| :|||: |||  
482 AC...GGCCGTGGCCTGGACTTCGGTGTCAACGTGCTGGTCAGCCAGAAG 528  
126 TyrGluPro 128  
|||||  
529 CCAGAGCCT 537

2

2

•  
•  
•  
•



OM of: US-09-135-238B-2\_COPY\_273\_390 to: Issued\_Patents\_NA.\* out\_format : pfs

Date: Sep 12, 2000 6:54 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=framet\_p2n.model -DEV=xlp  
-O=/cgn2\_1/USPTO.spool/US09135238/runat\_24082000\_091324\_17895/app\_query.fasta\_1.1282  
-DB=Issued\_Patents\_NA -QFMT=fastap -SUFFIX=p2n.rni  
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000  
-LOOPEXT=0.000 -GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000  
-XGAPEXT=0.500 -FGAPOP=6.000 -GAPEXT=7.000 -XGAPOP=10.000  
-YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1  
-MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200  
-THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=1000000  
-USER=US09135238 -CGN1\_1\_68 -NCPU=6 -ICPU=3 -LONGLOG -NO\_XLPXY  
-WAIT -THREADS=1

Search information block:

Query: US-09-135-238B-2\_COPY\_273\_390

Query length: 118

Database: Issued\_Patents\_NA.\*

Database sequences: 243080

Database length: 68777915

Search time (sec): 93.130000

score\_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-824-405-5	-	96.50	126.18	15.46	1683
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-264-861A-10	-	96.00	128.90	10.92	1150
/cgn2_6/ptodata/1/ina/PCRTUS_COMB.seq:PCT-US95-07784-10	-	96.00	128.90	10.92	1150
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-665-259-20	-	95.50	124.46	19.28	1743
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-482-677-9	-	94.00	122.35	25.26	1743
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-804-227C-7	-	94.00	93.31	1.0e+03	44377
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-804-198-1	-	94.00	93.31	1.0e+03	44377
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-482-385A-5	-	93.00	116.93	50.66	2728
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-469-412A-8	-	92.00	134.42	5.38	332
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-07-765-830A-5	-	91.00	110.12	121.29	4258
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-308-949A-1	-	89.50	108.44	150.49	4060
/cgn2_6/ptodata/1/ina/backfiles1.seq:3200327-7	-	89.00	128.59	11.35	397
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-318-193-17	-	89.00	128.50	11.48	401
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-318-193-11	-	89.00	125.92	15.99	535
/cgn2_6/ptodata/1/ina/backfiles1.seq:5200327-4	-	89.00	125.78	16.27	543
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-318-193-15	-	89.00	122.00	26.44	828
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-318-193-13	-	89.00	121.95	26.58	832
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-318-193-9	-	89.00	121.89	26.80	838
/cgn2_6/ptodata/1/ina/backfiles1.seq:5200327-5	-	89.00	121.81	27.10	846
/cgn2_6/ptodata/1/ina/backfiles1.seq:5200327-2	-	89.00	121.29	28.95	896
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-318-193-7	-	89.00	121.25	29.10	900
/cgn2_6/ptodata/1/ina/backfiles1.seq:5200327-3	-	89.00	121.20	29.28	905
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-318-193-9	-	89.00	121.16	29.43	909
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-405-175A-10	-	89.00	116.28	55.04	1566
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-428-415-1	-	89.00	112.38	90.77	2419
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-379-685-1	-	89.00	112.38	90.77	2419
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-854-029-1	-	89.00	112.38	90.77	2419
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-09-339-964-1	-	88.50	105.64	215.55	4743
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-459-041A-1	-	88.50	99.17	494.40	9759
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-999-733-1	-	88.50	99.17	494.40	9759
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-07-642-734C-3	-	88.50	92.62	1.1e+03	20235
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-989-363-1	-	88.00	123.91	20.69	572
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-264-526-1	-	88.00	123.91	20.69	572
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-152-483B-2	-	88.00	105.83	210.38	4294
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-713-118-1	-	87.50	100.41	421.70	7266
/cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-581-148C-30	-	87.00	100.92	394.72	6343
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-125-468-1	-	87.00	86.98	2.4e+03	30001
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-474-933-1	-	87.00	86.98	2.4e+03	30001
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-775-009-36	-	86.50	105.53	218.44	3507
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-455-543A-8	-	86.50	99.11	497.78	7175

/cgn2\_6/ptodata/1/ina/5C\_COMB.seq:US-08-193-078B-8 - 86.50 99.11 497.78 7175  
/cgn2\_6/ptodata/1/ina/5C\_COMB.seq:US-08-223-305D-8 - 86.50 99.11 497.78 7175  
/cgn2\_6/ptodata/1/ina/5C\_COMB.seq:US-08-149-097D-8 - 86.50 99.11 497.78 7175  
/cgn2\_6/ptodata/1/ina/5B\_COMB.seq:US-08-455-543A-7 - 86.50 98.88 512.73 7362  
/cgn2\_6/ptodata/1/ina/5C\_COMB.seq:US-08-193-078B-7 - 86.50 98.88 512.73 7362

seq\_name: /cgn2\_6/ptodata/1/ina/5D\_COMB.seq:US-08-824-405-5

seq\_documentation\_block:  
; Sequence 5, Application US/088244405  
; Patent No. 5962246  
; GENERAL INFORMATION:  
; APPLICANT: Ladner, Robert D.  
; APPLICANT: Lynch, Frank  
; APPLICANT: Caradonna, Salvatore J.  
; TITLE OF INVENTION: DUTPase, Its Isoforms, and  
; TITLE OF INVENTION: Diagnostic and Other Uses  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dechert, Price & Rhoads  
; STREET: 4006 Bell Atlantic Tower, 1717 Arch Stre  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19103-2793  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/824,405  
; FILING DATE: 26-MAR-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bloom, Allen  
; REGISTRATION NUMBER: 29,135  
; REFERENCE/DOCKET NUMBER: 350163-101  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 609/520-3214  
; TELEFAX: 609/520-3259  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1683 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-08-824-405-5

alignment\_scores:  
Quality: 96.50 Length: 86  
Ratio: 1.969 Gaps: 5  
Percent Similarity: 56.97 Percent Identity: 33.721

alignment\_block:

US-09-135-238B-2\_COPY\_273\_390 x US-08-824-405-5/rev ..  
Align seg 1/1 to reverse of: US-08-824-405-5 from: 1 to: 1683

22 ArgAlaLeuGluSerSerGlnArgProArgGlySerProArg..... 35

1020 AAAGCGCGCCCTCCAGGGGCCCTAATGGACCCACCCCGCCCGC 971

36 .ProArgSerGlnAsnAsn.....IleTyrSerAlaCysProA 48

970 CCCCCGGGGAGACACCCCGCTCCGCCCTTCTCCACAGCTCAA 921

48 rgArgAlaArgGlyAlaAspAlaAlaGlyThrGlyGluAlaPro...Val 63



```
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
PCT-US95-07784-10

alignment_scores:
    Quality: 96.00      Length: 137
    Ratio: 1.333       Gaps: 6
Percent Similarity: 52.555 Percent Identity: 29.927

alignment_block:
US-09-135-238B-2_COPY_273_390 x PCT-US95-07784-10 ..
Align seg 1/1 to: PCT-US95-07784-10 from: 1 to: 1150

1 LysArgAlaValGluArgArgLysAlaLeuSerArgArgAlaArgLe 17
697 CGTCGGAGGTCCACGACGTCGAGGCGGCTCGACGAGGTCTCGCGC 746
17 uAlaVal.....ArgMetArgAlaLeuGluSerSerGlnA 29
||||:||||:||||:||||:||||:||||:||||:||||:
747 TGCTCGCGCGCATGGCTTCACCGTCTGCGCGGAGGATCGCTGTT 796
29 rgProArgGlySerProArgProArgSerGlnAsnAlleTyrSerAla 45
||:||||:||||:||||:||||:||||:||||:||||:
797 GCGGCGAGGACATCCACGAGTCTGCGCGC.....ACGGCG 831
46 CysProArgArgAlaArgGlyAlaAspAlaAlaGlyThrGly..... 59
832 TCGGAGCACCTAGTACCGCGCGGCGGCTACCGCACCGCGGTGCGG 881
60 .....GluAlaProValPro.....GlyProGlyAlaP 69
882 TCGCGGGGCTCGCGACGTCGCGGAGTCTCTCGGAAGCTGCTGGCGGC 931
69 roLeuProProAlaProLeuGlnValSerGluSerProTTPLeuHisAla 85
||||:||||:||||:||||:||||:||||:||||:||||:
932 CTTACCGCGGAGTCTGCGGAACACGTTGAGGAGTCTGTTCCACGGT 981
86 ProSerLeuLysThrSerCysGluTyrValSerLeuTyrHisGlnPro.. 101
||||:||||:||||:||||:||||:||||:||||:||||:
982 CTGCGCGTGACGAACAGTCTCGCGGATCTCCCTGTTCTGACGCGGAC 1031
102 .....AlaAlaMetMetGluAspSerAspSer....AspAspTyrIleA 115
||||:||||:||||:||||:||||:||||:||||:||||:
1032 CGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1081
115 snValProAla 118
:::||||:
1082 GCTGCGCGGCGG 1092

seq_name: /cgn2_6/prodata/1/ina/6_COMB.seq.US-08-665-259-20

seq_documentation_block:
; Sequence 20, Application US/08665259
; Patent No. 6028173
; GENERAL INFORMATION:
; APPLICANT: Landes, Gregory M.
; APPLICANT: Burn, Timothy C.
; APPLICANT: Connors, Timothy D.
; APPLICANT: Dackowski, William R.
; APPLICANT: Van Raay, Terence J.
; APPLICANT: Klinger, Katherine W.
; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
; TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESS: GENZYME CORPORATION
; STREET: One Mountain Road
; CITY: Framingham
```

69 oLeuProProAlaProLeuGlnValSerGluSerProTriPheuHisAlap 86  
|||||  
1409 ACTGCCACCTGCACCG.....CATAGTC 1387

86 roSerLeuIysThrSer.....CysGluTyr 94

1386 CTTCGGCAGAACTCTTCAGGCTGATTCGGTAAC TGCCACGCCAGGTC 1337

95 valSerLeuTyrHisGlnProAlaAla 103

1336 TGC AATGTGACCTCACAGTCTCTGTGGCT 1310

name: /mnt6/ntdata3/1/ine/50 COMP ser:HS-08-804-227C-7

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

Sequence 7, Application US/08804227C

GENERAL INFORMATION:

APPLICANT: DEHOLL, Bradley S.  
APPLICANT: Kuhstoss, Stuart A.

APPLICANT: Rosteck, Paul R., Jr.  
APPLICANT: Sutton, Kimberly L.

TITLE OF INVENTION:	POLYKETIDE SYNTHASE GENES
NUMBER OF SEQUENCES:	15

ADDRESS: THOMAS C. PLANT 1501  
CORRESPONDENCE ADDRESS:

STREET: LILLY CORPORATE CENTER

STATE: IN

COUNTRY: USA  
ZIP: 46285

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS

SOFTWARE: ASCII(DOS) Text only  
CURRENT APPLICATION DATA.

APPLICATION NUMBER: US/08/804,227C

CLASSIFICATION: 435

NAME: Plant, Thomas, G.

REGISTRATION NUMBER: 35,784  
REFERENCE/DOCKET NUMBER: X-8231

**TELECOMMUNICATION INFORMATION:**  
**TELEPHONE: 317-276-2459**

INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS.

LENGTH: 44377 base pairs

STRANDEDNESS: single

MOLECULE TYPE: DNA (genomic)

FEATURE:	NAME/KEY:	CDS
----------	-----------	-----

LOCATION: 330..14002  
FEATURE:

NAME/KEY: CDS  
LOCATION: 14046..20036

FEATURE:	NAME / KEY:	CDS
----------	-------------	-----

LOCATION: 20110..31284

[illegible]

LOCATION: 31251:30074

FEATURE:

NAME/REF. CDS  
LOCATION: 36155..41830

08-804-227C-1

gment scores:

---

seq name: /cgn2 6/ptodata/1/ina/5p\_comb.seq:US-08-804-198-1

```

66 .....ProGlyAlaProL 70
27920 CCGTGGTAGAGTCGCCGATGTCAGTTGTCGGCGCGGGGGCGGGGT 27871

70 euProPro.....Ala 73
|||||
27870 GCGCGCGCGAGCGTCGCGCGCGTGTCTGTCTGCGCGGTGAGCAGCC 27821

74 ProLeuGlnValSerGluSerProTip 82
|||||
27820 CCGCGCGTGCAGGTCACGTCCTGG 27794

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-482-385A-5

seq_documentation_block:
; Sequence 5, Application US/08482385A
; Patent No. 5728561
; GENERAL INFORMATION:
; APPLICANT: DENOVA, CLAUDIO D.
; TITLE OF INVENTION: GENES ENCODING BRACHED CHAIN ALPHA
; TITLE OF INVENTION: KETOACID DEHYDROGENASE FROM STREPTOMYCES SPECIES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETER C. RICHARDSON
; STREET: 235 EAST 42ND STREET, 20TH FLOOR
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A
; ZIP: 10017-5755
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/482.385A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SHEYKA, ROBERT F.
; REGISTRATION NUMBER: 31,304
; REFERENCE/DOCKET NUMBER: PC8346C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-573-1189
; TELEFAX: 212-573-1939
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2728 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-482-385A-5

alignment_scores:
Quality: 93.00 Length: 106
Ratio: 1.691 Gaps: 7
Percent Similarity: 51.887 Percent Identity: 34.906

alignment_block:
US-09-135-238B-2_COPY_273_390 x US-08-482-385A-5 ..

Align seg 1/1 to: US-08-482-385A-5 from: 1 to: 2728

1 LysArgAlaValGluArgLysAlaLeuSerArg..... 13
:|||||:
3 CGACGCGGCTCCGAAACCGCGATCACCGCTGTCGTCGATGAGCGGGTTG 52
14 ....AlaArg.....LeuAlaValArgMetArgAlaLeuGluSerS 27
:|||||:
53 ATTCGCGGTAGGATGTCGACGCGACACGCGCGCTCGCGCACCGGA 102

```

```

27 erGlnArgProArgGlySerProArgSerGlnAsnAsnIleTyr 43
:|||||:
103 CCGTATCGAGCGCGCGCTCCGCTGGAGCATCT..... 137

44 SerAlaCysProArgArgAlaArg.....GlyAlaaspAlaGlyTh 58
:|||||:
138 GGAGGATGTCCTGATGATGCGCTGCCAGCGCGCGCGCGCGAGGGT 187

58 rGlyGluAlaProValProGlyPro.....GlyAlaProLeuProProA 73
:|||||:
188 ACCCGGGCTCCGCTCCGCGCATTTGTCAGTGCATGCTCTCCGG 237

73 lacProLeuGlnValSerGluSerProTrp.....LeuHisAlaPro 86
:|||||:
238 CCTCCTTA.....CCATGGACGTAGTGGTTTCATTCACG 272

87 SerLeuLysThrSerCys 92
:|||||:
273 CTGTGGAGAACCGTTGT 290

seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-469-412A-8

seq_documentation_block:
; Sequence 8, Application US/08469412A
; Patent No. 5856125
; GENERAL INFORMATION:
; APPLICANT: Mavrothalassitis, George J.
; APPLICANT: Blair, Donald G.
; APPLICANT: Fisher, Robert J.
; APPLICANT: Beal Jr., Gregory J.
; APPLICANT: Athanasios, Metropi A.
; APPLICANT: Scuras, Dionysios N.
; TITLE OF INVENTION: The ERF Genetic Locus and Its Products
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469.412A
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Garrett-Wackowski, Eugenia
; REGISTRATION NUMBER: 37,330
; REFERENCE/DOCKET NUMBER: 015280-229000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 332 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..332
; OTHER INFORMATION: /note= "human ERF gene promoter region"
US-08-469-412A-8

alignment_scores:

```

```

; MOLECULE TYPE: DNA (genomic)
US-07-765-830A-5

alignment_scores:
  Quality: 91.00 Length: 122
  Ratio: 1.596 Gaps: 5
  Percent Similarity: 46.721 Percent Identity: 26.230

alignment_block:
US-09-135-238B-2_COPY_273_390 x US-07-765-830A-5/rev ..
Align seg 1/1 to reverse of: US-07-765-830A-5 from: 1 to: 4258

      8 LysAlaLeuSerArgArgAlaArgAlaValArgMetArgAlaLe 24
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2452 CGCGCGTTGGGGTGCTCTTGAGAGAGCGGCCACCGCTGCCCGGACTT 2403

      24 uGluSerSerGlnArgProGlySer..... 33
      | |||:|||||:|||||:|||||:|||||:|||||:|||||:
2402 GGTGTCCACGGCAGGTCCCGGAGCAGTCGCGACCGGTGCGCCTTGAGAT 2353

      34 .....ProArgPro 36
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2352 TGGCGCCCCCGCCCCCGGAGCCTTGTGCGCCTTCTTGTGACCGCGCCC 2303

      37 .....ArgSerGlnAsnAsnIleTyrSerAlaCy 46
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2302 GCAGCGTGGGCTCGGCAGCTCTCTGCGCGGGGTTCGGGGACCTG 2253

      46 sProArgArgAlaArg.....GlyAlaAspAlaAlaGlyThrGlyG 60
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2252 TCCGAGGAAGAGCGCGCAGGTGAAGGACACCGCGCTGCAGCAGGGG 2203

      60 luAlaProValProGlyProGlyAla..... 68
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2202 ACTGTGATGCTCCAGCCCCACCGCGCGCGCGGGGTGTCCTCCCA 2153

      69 ...ProLeuProAlaProLeuGlnValSerGluSerProTyrLeuHi 84
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2152 AGCCAGAGCGCGCCCCACCGATGCGCGGCGACGTTGGCCCG.....CA 2109

      84 sAlaProSerLeuLys 89
      | |||:|||||:|||||:|||||:|||||:|||||:|||||:
2108 TCACCTGCCACGCGG 2093

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-308-949A-1

seq_documentation_block:
; Sequence 1, Application US/08308949A
; Patent No. 5580703
; GENERAL INFORMATION:
; APPLICANT: Kotin, Robert M.
; APPLICANT: Berns, Kenneth I.
; APPLICANT: Linden, Ralph M.
; TITLE OF INVENTION: Human Adeno-Associated Virus Integration
; TITLE OF INVENTION: Site DNA and Uses Thereof
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07765,830A
; FILING DATE: 19910926
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SCOTT, WATSON T
; REGISTRATION NUMBER: 26581
; REFERENCE/DOCKET NUMBER: 5683/92723
; TELEPHONE: 202-861-3067
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4258 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
```

```

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/947,127
; FILING DATE: September 27, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Carroll, Alice O.
; REGISTRATION NUMBER: 33,542
; REFERENCE/DOCKET NUMBER: ACC92-10F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4060 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-308-949A-1

alignment_scores:
  Quality: 89.50      Length: 125
  Ratio: 1.467        Gaps: 6
  Percent Similarity: 48.800  Percent Identity: 31.200

alignment_block:
US-09-135-238B-2_COPY_273_390 x US-08-308-949A-1 ..
Align seg 1/1 to: US-08-308-949A-1 from: 1 to: 4060

1 LysArgAlaValGluArgArgLysAlaLeuSerArgAlaArgLe 17
  :||||| :||||| :||||| :||||| :|||||
528 CGCGGAGCTGGCCCGGAGAGCGCGCGCGCGCGCGCGCG 577
  :||||| :||||| :||||| :||||| :|||||
17 uAlaValArgMet.....ArgAlaL 24
  :||||| :||||| :||||| :||||| :|||||
578 GCGCGCGGAGCTTCGCGCGCGCTGTGCGGCGCGGAGCTG 627
  :||||| :||||| :||||| :||||| :|||||
24 euGluSerSerGlnArgProArgGlySerProArgSerGln 40
  :||||| :||||| :||||| :||||| :|||||
628 CGTCTGATGCTGGCGCGCGCGCGCGCGCGCGCGCGCG 677
  :||||| :||||| :||||| :||||| :|||||
41 AsnIleTyrSerAlaCysProArgAla.....ArgG 52
  :||||| :||||| :||||| :||||| :|||||
678 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGTGCGGAG 727
  :||||| :||||| :||||| :||||| :|||||
52 yAlaAspAlaAlaGlyThrGlyGluAlaPro..... 66
  :||||| :||||| :||||| :||||| :|||||
728 GTATCAGCGCGCTGCACAGGTGACGCGCGCGCGCGCG 777
  :||||| :||||| :||||| :||||| :|||||
66 roGlyAla.ProLeu.....ProProAlaProLeuGln 77
  :||||| :||||| :||||| :||||| :|||||
778 CAGGTCACCTCTTCGCCACCTGGGGGATCTCTCCCTCC 827
  :||||| :||||| :||||| :||||| :|||||
77 lserGluSerProTrpLeuHisAlaProSerLeuLysThr 93
  :||||| :||||| :||||| :||||| :|||||
828 CTCGATCGCGCGCGCTGTTACTGGCGCTTNCACCTATG 877
  :||||| :||||| :||||| :||||| :|||||
94 .....TyrValSerLeuTyrHis 99
  :||||| :||||| :||||| :||||| :|||||
878 CCGCGTTCAGTCCCTCCCTTACCAT 900

seq_name: /cgn2_6/ptodata/1/ina/backfiles1.seq:5200327-7

seq_documentation_block:
; Patent No. 5200327
; APPLICANT: GARVIN, ROBERT T.; MALEK, LAWRENCE T.
; TITLE OF INVENTION: EXPRESSION SYSTEM FOR THE SECRETION OF
; BIOACTIVE HUMAN GRANULOCYTE MACROPHAGE COLONY STIMULATING
; FACTOR (GM-CSF) AND OTHER HETEROLOGOUS PROTEINS FROM
; STREPTOMYCES
; NUMBER OF SEQUENCES: 24

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/224,568
; FILING DATE: 26-JUL-1988
; SEQ ID NO:7:
; LENGTH: 397
; 5200327-7

alignment_scores:
  Quality: 89.00      Length: 96
  Ratio: 1.745        Gaps: 7
  Percent Similarity: 53.125  Percent Identity: 37.500

alignment_block:
US-09-135-238B-2_COPY_273_390 x 5200327-7 ..
Align seg 1/1 to: 5200327-7 from: 1 to: 397

1 LysArgAlaValGluArgArgLysAlaLeuSerArgAlaArgLe 17
  :||||| :||||| :||||| :||||| :|||||
66 CGACGCCCATGTGACCGCTGCTGCTGCTGCTGCTGCTGCT 115
  :||||| :||||| :||||| :||||| :|||||
17 u.....Ala.ValArgMetArgAlaLeuGluSerSer.G 31
  :||||| :||||| :||||| :||||| :|||||
116 CGCGGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCG 163
  :||||| :||||| :||||| :||||| :|||||
31 gGlySerProArgSerGlnAsnAsnIle..... 42
  :||||| :||||| :||||| :||||| :|||||
164 .GGAGGCGTGGCGCTGCGCGGAGACCGCGCTGCTGCTG 212
  :||||| :||||| :||||| :||||| :|||||
43 .....TyrSerAlaCysProArgArgAlaArgGlyAla 56
  :||||| :||||| :||||| :||||| :|||||
213 AGCGGTACCTCTGCGCTGCGCGGCGCTCGACCGCGCG 262
  :||||| :||||| :||||| :||||| :|||||
57 GlyThrGlyGluAla.ProValProGly.....ProGly 72
  :||||| :||||| :||||| :||||| :|||||
263 GGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 312
  :||||| :||||| :||||| :||||| :|||||
72 roAlaProLeuGlnValSerGluSerProTrp 82
  :||||| :||||| :||||| :||||| :|||||
313 CCACCGCGCGCACAGAATGTCGGAACCCCTAC 344

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-318-193-17

seq_documentation_block:
; Sequence 17, Application US/08318193
; Patent No. 5641663
; GENERAL INFORMATION:
; APPLICANT: GARVIN, ROBERT T.
; APPLICANT: MALEK, LAWRENCE T.
; TITLE OF INVENTION: AN EXPRESSION SYSTEM FOR THE SECRETION
; OF BIOACTIVE HUMAN GRANULOCYTE MACROPHAGE COLONY
; STIMULATING FACTOR (GM-CSF) AND OTHER HETEROLOGOUS
; PROTEINS FROM STREPTOMYCES
; NUMBER OF SEQUENCES: 91
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/08/318,193
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,314

```



```

; FILING DATE:
; APPLICATION NUMBER: US 07/224,568
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 18740/116 CACO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 401 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Other nucleic acid;
; DESCRIPTION: Synthetic DNA oligonucleotide
US-08-318-193-17

alignment_scores:
    Quality: 89.00      Length: 96
    Ratio: 1.745       Gaps: 7
Percent Similarity: 53.125 Percent Identity: 37.500

alignment_block:
US-09-135-238B-2_COPY_273_390 x US-08-318-193-17 ..
Align seg 1/1 to: US-08-318-193-17 from: 1 to: 401

1 LysArgAlaValGluArgArgLysAlaLeuSerArgAlaArgLe 17
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
66 CGAGCGCCATGTGACCGCTGCTGCTGCGCGCGCGCGCGAGCT 115
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
17 u.....Ala.ValArgMetArgAlaLeuGluSerSer.GlnArgProAr 31
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
116 CGCGGGGGCGGACCGCGCGCGCGCGCGAGGTCTCGCGCGCGACGG.. 163
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
31 gGlySerProArgProArgSerGlnAsnAsnIle..... 42
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
164 .GGAGCGGTGCGCGCGCGCGCGAGACCGCGCTGCTGCGCGCTCACGG 212
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
43 .....TyrSerAlaCysProArgArgAlaArgGlyAlaAspAlaA 56
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
213 AGCGGTACCTCTCGCCCTCGCGCGCGCGCGCTGCTGCGCGCTCACGG 262
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
57 GlyThrGlyGluAla.ProValProGly...ProGlyAlaProLeuProP 72
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
263 GGCACCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 312
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
72 roAlaProLeuGlnValSerGluSerProTrp 82
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
313 CCCACCGCGCACAGAAGATGTCGGAACCCCTAC 344

```

seq\_name: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:US-08-318-193-11

```

seq_documentation_block:
; Sequence 11, Application US/08318193
; Patent No. 5641663
; GENERAL INFORMATION:
; APPLICANT: GARVIN, Robert T.
; APPLICANT: MALEK, Lawrence T.
; TITLE OF INVENTION: AN EXPRESSION SYSTEM FOR THE SECRETION
; TITLE OF INVENTION: OF BIOACTIVE HUMAN GRANULOCYTE MACROPHAGE COLONY
; TITLE OF INVENTION: STIMULATING FACTOR (GM-CSF) AND OTHER HETEROLOGOUS
; TITLE OF INVENTION: PROTEINS FROM STREPTOMYCES
; NUMBER OF SEQUENCES: 91
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: Virginia

```

```

; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/318,193
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,314
; FILING DATE:
; APPLICATION NUMBER: US 07/224,568
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 18740/116 CACO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 535 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Other nucleic acid;
; DESCRIPTION: Synthetic DNA oligonucleotide
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 399..533
US-08-318-193-11

```

```

alignment_scores:
    Quality: 89.00      Length: 96
    Ratio: 1.745       Gaps: 7
Percent Similarity: 53.125 Percent Identity: 37.500

alignment_block:
US-09-135-238B-2_COPY_273_390 x US-08-318-193-11 ..
Align seg 1/1 to: US-08-318-193-11 from: 1 to: 535

```

```

1 LysArgAlaValGluArgArgLysAlaLeuSerArgAlaArgLe 17
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
66 CGAGCGCCATGTGACCGCTGCTGCTGCGCGCGCGCGCGAGGTCT 115
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
17 u.....Ala.ValArgMetArgAlaLeuGluSerSer.GlnArgProAr 31
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
116 CGCGGGGGCGGACCGCGCGCGCGCGCGAGGTCTGCTGCGCGCGACGG.. 163
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
31 gGlySerProArgProArgSerGlnAsnAsnIle..... 42
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
164 .GGAGCGGTGCGCGCTCGCGCGGAGACCGCGCGCTGCTGCTGCGCGCTCACGG 212
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
43 .....TyrSerAlaCysProArgArgAlaArgGlyAlaAspAlaA 56
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
213 AGCGGTACCTCTCGCCCTCGCGCGCGCGCGCTGCGCGCGCGGACCTCC 262
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
57 GlyThrGlyGluAla.ProValProGly...ProGlyAlaProLeuProP 72
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
263 GGCACCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 312
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
72 roAlaProLeuGlnValSerGluSerProTrp 82
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
313 CCCACCGCGCACAGAAGATGTCGGAACCCCTAC 344

```

seq\_name: /cgn2\_6/ptodata/1/ina/backfiles1.seq:5200327-4

```
seq_documentation_block:
; Patent NO. 5200327
; APPLICANT: GARVIN, ROBERT T.; MALEK, LAWRENCE T.
; TITLE OF INVENTION: EXPRESSION SYSTEM FOR THE SECRETION OF
; BIOACTIVE HUMAN GRANULOCYTE MACROPHAGE COLONY STIMULATING
; FACTOR (GM-CSF) AND OTHER HETEROLOGOUS PROTEINS FROM
; STREPTOMYCES
; NUMBER OF SEQUENCES: 24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/224,568
; FILING DATE: 26-JUL-1988
; SEQ ID NO: 4:
; LENGTH: 543
5200327-4
```

```
alignment_scores:
Quality: 89.00 Length: 96
Ratio: 1.745 Gaps: 7
Percent Similarity: 53.125 Percent Identity: 37.500
```

## alignment\_block:

US-09-135-238B-2\_COPY\_273\_390 x 5200327-4 ..

Align seg 1/1 to: 5200327-4 from: 1 to: 543

```
1 LysArgAlaValGluArgLysAlaLeuSerArgAlaArgLe 17
:::|||||:::|||||:::|||||
66 CGACGCCCATGTGACCGCTGCGTGTCTGCGGGCGCGCGAGGCT 115
17 u.....Ala.ValArgMetArgAlaLeuGluSerSer.GlnArgProAr 31
|||:::|||||:::|||||:::|||||
116 CGCGGGGGCGGACCGGACCGCGCGGAGGTCTCGCGCGGACCG.. 163
31 gGlySerProArgProArgSerGlnAsnAsnIle..... 42
|||:::|||||:::|||||:::|||||
164 .GGAGCGTGGCGCTCGCGGAGAGACCGCGTCTGCTGCGGCTCAGG 212
43 .....TyrSerAlaCysProArgArgAlaArgGlyAlaAspAlaAla 56
|||:::|||||:::|||||
213 AGCGGTACCTCTCGCTGCGGCGGCGCTCGACCGCGCGGACCTCC 262
57 GlyThrGlyGluAla.ProValProGly...ProGlyAlaProLeuProP 72
|||||:::|||||:::|||||:::|||||
263 GGCACGGGGCGCGGCGGACCGCGGCGGACCGGTCGCGGCGGCGGCCC 312
72 roAlaProLeuGlnValSerGluSerProTrp 82
|||:::|||||
313 CCCACCGGCACAGAATGTCCGGAACCTAC 344
```

OM of: US-09-135-238B-2 to: GenEmbl.\* out\_format : pfs  
Date: Sep 12, 2000 5:34 AM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:  
-MODE=frame+p2n.model -DEV=xlpx  
-Q/cn2\_1/QPMT=spool/US09135238/runat\_24082000\_091113\_16775/app\_query.fasta\_1.i282  
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-GAPOP=6.000 -GAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=biosum62  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NRM=ext -MINLEN=0 -MAXLEN=1000000  
-USER=US09135238 -CGN1\_1\_5120 -NCPU=6 -ICPU=3 -LONGLOG -NO\_XLPPX  
-WAIT -THREADS=1

Search information block:  
Query: US-09-135-238B-2  
Query length: 390  
Database: GenEmbl.\*  
Database sequences: 972840  
Database length: 892348106  
Search time (sec): 2785.000000

score_list:	Strd Orig	EScore	Len	Documentation
Sequence	+ 2047.00	1733.42	1339	! AF057557 Homo sapiens anti-Fas
gb_pr3:AF057557	+ 661.50	528.40	251	! AC023534 Homo sapiens clone
gb_btgl1:AC023534	- 284.00	220.68	0.0003	! AC027719 Homo sapiens chromos
gb_btg22:AC027719	+ 198.00	170.58	0.2098	! U06431 Mus musculus B6/CBAF1J
gb_ro:MMU06431	+ 192.00	166.16	0.3699	! AF091137 Trichosurus vulpecula
gb_om:AF091137	+ 188.00	177.35	0.0880	! U83428 Mus musculus polymeric i
gb_ro:MMP1GR03	+ 188.00	175.90	0.1061	! Y16525 Mus musculus pigr gene,
gb_ro:MMP1GR3	+ 188.00	145.10	5.51	! E15253 Mouse pig receptor gen
gb_pat:E15253	+ 188.00	145.10	5.51	! AB001489 Mus musculus DNA for
gb_ro:AB001489	+ 187.50	165.39	0.4080	! X15470 cDNA encoding immunity
gb_pat:E15470	+ 185.50	159.64	0.8532	! X15741 Rat mRNA for polymeric
gb_ro:RNIGR	+ 184.00	160.53	0.7613	! AB032195 Sus scrofa mRNA for p
gb_om:AB032195	+ 183.00	156.96	1.20	! L04797 Bos taurus polymeric in
gb_om:BOVP1GR	+ 182.00	173.73	0.1400	! S43441 transmembrane secretory
gb_pr2:S43435503	+ 182.00	160.66	0.7483	! A52094 Sequence 4 from Patent
gb_pat:A52094	+ 182.00	160.30	0.7839	! A52093 Sequence 3 from Patent
gb_ro:A52093	+ 182.00	159.02	0.9239	! A52091 Sequence 1 from Patent
gb_pat:A52091	+ 182.00	157.61	1.11	! S62403 transmembrane secretory
gb_pr2:S62403	+ 182.00	157.61	1.11	! X73079 Homo sapiens encoding F
gb_pr1:HSP1R	+ 182.00	137.55	1.12	! X73079 Homo sapiens encoding F
gb_om:BTPIGR1	+ 182.00	145.90	1.38	! X81371 B.taurus mRNA for bovin
gb_ro:BTPIGR1	+ 171.00	149.35	3.19	! AF035963 Rattus norvegicus kid
gb_ro:AF035963	- 167.00	112.12	378.14	! AC023534 Homo sapiens clone
gb_btgl1:AC023534	+ 166.50	143.09	7.13	! X00412 Rabbit mRNA for poly-im
gb_om:OCIGRECR	- 159.50	118.29	171.52	! AL109732 Streptomyces coelic
gb_btgl1:SC7H2	+ 157.00	137.60	14.41	! M24559 Human poly-Ig receptor
gb_pr1:HUM1GRPOLY	+ 152.50	146.79	4.44	! Y16527 mus musculus pigr gene,
gb_ro:MMP1GR5	+ 152.50	141.07	9.23	! A7238323 Homo sapiens mRNA for
gb_pr2:HSAX38323	+ 152.50	141.07	9.24	! AJ010101 Homo sapiens mRNA for
gb_pr2:HSAX010101	+ 152.50	136.15	17.37	! A7224864 Homo sapiens mRNA for
gb_pr2:HSAX224864	+ 147.50	108.02	640.04	! U15177 Human cosmid CRI-JC201
gb_pr2:HSU15177	+ 145.00	94.11	3.8e+03	! AL157395 Homo sapiens chrom
gb_btgl2:AL157395	+ 143.00	138.38	13.03	! U83430 Mus musculus polymeric i
gb_ro:MMP1GR05	- 142.50	102.95	1.2e+03	! AC020151 Drosophila melanog
gb_btgl6:AC020151	- 142.50	89.56	6.8e+03	! AE003551 Drosophila melanog
gb_in2:AE003551	+ 142.00	100.37	1.7e+03	! AB032367 Streptomyces avermit
gb_btgl1:AB032367	- 141.50	110.03	494.93	! AE001986 Deinococcus radiodur
gb_bt2:AE001986	- 140.00	102.68	1.3e+03	! AL031910 Leishmania major Fri
gb_in1:LMFL2743	+ 140.00	101.40	1.5e+03	! AL078610 Streptomyces coelic
gb_btgl1:SCH35	+ 140.00	91.53	5.3e+03	! AC010208 Homo sapiens clone
gb_btgl9:AC010208	- 140.00	90.07	6.4e+03	! AC009813 Homo sapiens chrom
gb_btgl26:AC009813	- 140.00	90.07	6.4e+03	! AC009813 Homo sapiens chrom

gb\_ba2:AF047717 + 139.00 113.87 302.25 7836 ! AF047717 Streptomyces chrys  
gb\_ba1:SC9B5 - 138.50 101.82 1.4e+03 36368 ! AL035206 Streptomyces coel  
gb\_ov:GDCEBP + 137.50 124.25 79.84 1686 ! X6844 G-domesticus gene fo  
gb\_in2:AF227923 + 137.50 120.73 125.41 2683 ! AF227923 Tribolium castaneu  
gb\_ba1:DVAPSAB + 137.50 118.98 156.92 3379 ! Z69372 D.vulgaris apsA and  
seq\_name: gb\_pr3:AF057557  
seq\_documentation\_block:  
LOCUS AF057557 1339 bp mRNA PRI 04-JUN-1998  
DEFINITION Homo sapiens anti-Fas-induced apoptosis (TOSO) mRNA, complete cds.  
ACCESSION AF057557  
VERSION AF057557.1 GI:3169292  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 1339)  
Hitoshi Y., Lorenz J., Kitada, S.I., Fisher, J., LaBarge, M.,  
Ring, H.Z., Francke, U., Reed, J.C., Kinoshita, S. and Nolan, G.P.  
Toso, a cell surface, specific regulator of Fas-induced apoptosis  
in T cells  
JOURNAL Immunology 8 (4), 461-471 (1998)  
MEDLINE 98246048  
REFERENCE 2 (bases 1 to 1339)  
LaBarge, M. and Hitoshi, Y.  
Direct Submission  
AUTHORS  
TITLE  
JOURNAL Submitted (03-APR-1998) Molecular Pharmacology, Stanford  
University, 300 Pasteur Dr., Palo Alto, CA 94304, USA  
FEATURES  
Location/Qualifiers  
1..1339  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1q31-q32"  
/cell\_line="Jurkat"  
/cell\_type="T cell"  
/note="TOSO.2 cDNA: cloned by PCR using primers designed  
from the original TOSO.1 cDNA sequence"  
1..1339  
/gene="TOSO"  
20..1192  
/gene="TOSO"  
/note="transmembrane protein; Ig superfamily member"  
/codon\_start=1  
/product="anti-Fas-induced apoptosis"  
/protein\_id="AAC18830.1"  
/db\_xref="GI:3169293"  
/translation="MDFWLPYLPVSGALRILPEYKVEGELSGSVTIKCPLEMHV  
RIVLCREMAGSGTCGVVSTNTEIKAEYKGRVTLKQYPRKNLFLVEVFTOLTESDSGVY  
ACGAMTDRGKTOKTVLNHSEYEPSEKQPMETPKWHLPLYFOMPAVASSKFEV  
TRVTPAQGRKVPVPHSSPTQIHRPRVSRASVAGDAPRFLPSTASKISALEG  
LLKPTQSYNHTLHRQRALDYQSGREGQGHILITLGLFLALLGLLVVKKRAV  
ERRKLRRLRAVRRALESQSRGPRSPRNNIYSACPRRAGDAAGTGEAP  
VPGGAPLPAPLQVSESPWLHAPSLKTSCEYSLYHQPAAMMEDSDSDYINVPA"  
127  
/gene="TOSO"  
/note="occurs in TOSO.1 cDNA"  
/replace="a"  
1015  
/gene="TOSO"  
/note="occurs in TOSO.1 cDNA"  
/replace="g"  
309 a 432 c 340 g 258 t  
BASE COUNT  
ORIGIN  
alignment\_scores:  
Quality: 2047.00 Length: 390  
Ratio: 5.262 Gaps: 0  
Percent similarity: 99.744 Percent Identity: 99.744

alignment\_block:

US-09-135-238B-2 x AF057557 ..

Align seg 1/1 to: AF057557 from: 1 to: 1339

1 MetAspArgTrrPLeuTrrProLeuTrrPheLeuProValSerGlyAlaLe 17  
|||||  
20 ATGGACCTTCCTGGCTTTGGCCACTTTACTTCTGCCAGTATCGGGGGCCCT 69  
17 uArgileLeuProGluValLysValGluGlyGlySerValT 34  
|||||  
70 GAGGATCTCTCCACAGTAAGGTAGAGGGGAGCTGGCGGATCAGTTA 119  
34 hrIleLysCysProLeuProGluMetHisValArgIleTrrLeuCysArg 50  
|||||  
120 CCATCAAGTGCCTTCTCTGAAATGCATGTGAGGATATATCTGTGCGG 169  
51 GluMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsnPh 67  
|||||  
170 GAGATGGCTGGATCTGGACATGTGTGTACCGTGTATCCACCACCACTT 219  
67 elleLysAlaGluTrrLysGlyArgValThrLeuLysGlnTrrProArgL 84  
|||||  
220 CATCAAGCAGATACAAAGGGCGGAGTTACTCTGAAGCAATACCCACGCA 269  
84 ysAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerGly 100  
|||||  
270 AGAATCTGTTCTAGTGGAGGTAAACACAGCTGACAGAAAGTGACAGCGGA 319  
101 ValTrrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLy 117  
|||||  
320 GTCTATGCTCGGAGCGGCATGAACACAGACGGGGAAAGACCCAGAA 369  
117 svAlThrLeuAsnValHisSerGluTrrGluProSerTrrPglGluGlnP 134  
|||||  
370 AGTCACCTGAACTGTCACAGTGAATACGAGCCATCATGGGAAGAGCAGC 419  
134 roMetProGluThrProLysTrrPheHisLeuProTrrLeuPheGlnMet 150  
|||||  
420 CAATGCCTAGACTCCAAAATGGTTTCATCTGCCCTATTTGTTCCAGATG 469  
151 ProAlaTrrAlaSerSerLysPheValThrArgValThrThrProAl 167  
|||||  
470 CTGCTATGCTAGCTTCTTCCAAATTCGTAAACAGAGTTACACACCCAGC 519  
167 aGlnArgGlyLysValProProValHisHisSerSerProThrThrGlnI 184  
|||||  
520 TCAAAGGGGCAAGTCCCTCCAGTTCACCACTCTCTCCCCACCCACCAAA 569  
184 leThrHisArgProArgValSerArgAlaSerSerValAlaGlyAsPlys 200  
|||||  
570 TCACCCACCGCCCTCCAGTGTCCAGAGCATCTTCAGTAGCAGGTGACAAG 619  
201 ProArgThrPheLeuProSerThrThrAlaSerLysIleSerAlaLeuGl 217  
|||||  
620 CCCCAGAACCTTCCTGCCATCCACTACAGCTCAAAATCTCAGCTCTGGA 669  
217 uGlyLeuLeuLysProGlnThrProSerTrrAsnHisHisThrArgLeuH 234  
|||||  
670 GGGGTGCTCAAGCCCGACGCCAGCTACAAACCCACACACCCAGGCTGC 719  
234 lsArgGlnArgAlaLeuAspTrrGlySerGlnSerGlyArgGluGlyGln 250  
|||||  
720 ACAGGCAGACAGACACTGGACTATGGCTACAGTCTGGGAGGAAGGCCAA 769  
251 GlyPheHisIleLeuIleProThrIleLeuGlyLeuPheLeuLeuAlaLe 267  
|||||  
770 GGATTTCCATCTGATCCGACCATCTCTGGGCTTTTCTCTGCTGGCACT 819  
267 uleuGlyLeuValValLysArgAlaValGluArgArgLysAlaLeuSerA 284  
|||||  
820 TCTGGGCTGCTGGTGAAGGGCGGTGAAGAGGAGGAAGCCCTCTCCA 869

284 tGArgAlaArgGLeuAlaValArgMetArgAlaLeuGluSerSerGln 300  
|||||  
870 GGCGGGCCCGCCGACTGGCCGTGAGATGGCGCCCTGGAGAGTCCCCAG 919  
301 ArgProArgGlySerProArgProArgSerGlnAsnAsnIleTrrSerAl 317  
|||||  
920 AGGCCCGCGGGTGCCTGCCGACCGCTCCCCAAAACACATCTACAGCCG 969  
317 acYsProArgArgAlaArgGlyAlaAspAlaAlaGlyThrGlyGluAlaP 334  
|||||  
970 CTGCCCGCGCGCGCTCTGTGAGCGGACGCTCGAGGCACAGGGGAAAGCC 1019  
334 roValProGlyProGlyAlaProLeuProProAlaProLeuGlnValSer 350  
|||||  
1020 CCGTTCCTCCCGCCGAGCGCGCTGCCCCCGCCCGCTGCAGGTGTCT 1069  
351 GluSerProTrrPheHisAlaProSerLeuLysThrSerCysGluTrrVa 367  
|||||  
1070 GAATCTCCCTGGCTCCATGCTGAGCGGACGCTCTGAAGACCACTGTGAATACCT 1119  
367 lSerLeuTrrHisGlnProAlaAlaMetMetGluAspSerAspSerAspA 384  
|||||  
1120 GAGCTCTACCAACAGCTGCGCCATGATGGAGGACAGTGATTCAGATG 1169  
384 spTrrIleAsnValProAla 390  
|||||  
1170 ACTACATCAATGTTCTCTGCC 1189  
seq\_name: gb\_htg11:AC023534  
seq\_documentation\_block:  
LOCUS AC023534.2 221365 bp DNA HTG 02-MAR-2000  
DEFINITION Homo sapiens clone RP11-462N18, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 61  
unordered pieces.  
ACCESSION AC023534.2 GI:7143452  
VERSION AC023534.2  
KEYWORDS HTGS\_PHASE1.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 221365)  
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
TITLE Homo sapiens chromosome, clone RP11-462N18  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 221365)  
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Bedalov,F., Boguslavsky,L.,  
Boukhvalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,  
Choepe,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,  
Dearellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,  
Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,  
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
Klein,J., Landers,T., Laroque,K., Lechoczky,J., Levine,R.,  
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,  
McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrum,J.,  
Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,  
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivari,T.M.,  
Peterson,K., Pierre,N., Pisanic,C., Pollara,V., Raymond,C.,  
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,  
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,  
Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,  
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and  
Zody,M.  
Direct Submission  
TITLE  
JOURNAL  
COMMENT Submitted (15-FEB-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 2, 2000 this sequence version replaced gi:6978210.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WtBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Project name: L5244  
 Center clone name: 462\_N\_18

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 61 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```

1 1108: contig of 1108 bp in length
1109 1208: gap of 100 bp
1209 2292: contig of 1084 bp in length
2293 2392: gap of 100 bp
2393 3673: contig of 1281 bp in length
3674 3773: gap of 100 bp
3774 4947: contig of 1174 bp in length
4948 5047: gap of 100 bp
5048 6104: contig of 1057 bp in length
6105 6204: gap of 100 bp
6205 7259: contig of 1055 bp in length
7260 7359: gap of 100 bp
7360 8667: contig of 1308 bp in length
8668 8767: gap of 100 bp
8768 10566: contig of 1799 bp in length
10567 10666: gap of 100 bp
10667 12173: contig of 1507 bp in length
12174 12273: gap of 100 bp
12274 13643: contig of 1370 bp in length
13644 13743: gap of 100 bp
13744 15215: contig of 1472 bp in length
15216 15315: gap of 100 bp
15316 16603: contig of 1294 bp in length
16610 16709: gap of 100 bp
16710 18085: contig of 1376 bp in length
18086 18185: gap of 100 bp
18186 19668: contig of 1483 bp in length
19669 19768: gap of 100 bp
19769 22083: contig of 2315 bp in length
22084 22183: gap of 100 bp
22184 24052: contig of 1869 bp in length
24053 24152: gap of 100 bp
24153 25902: contig of 1750 bp in length
25903 26002: gap of 100 bp
26003 27639: contig of 1637 bp in length
27640 27739: gap of 100 bp
27740 29377: contig of 1638 bp in length
29378 29477: gap of 100 bp
29478 30910: contig of 1433 bp in length
30911 31010: gap of 100 bp
31011 31225: contig of 215 bp in length
31226 31325: gap of 100 bp
31326 32890: contig of 1565 bp in length
32891 32990: gap of 100 bp
32991 34812: contig of 1822 bp in length
34813 34912: gap of 100 bp
34913 36817: contig of 1905 bp in length
36818 36917: gap of 100 bp
36918 38575: contig of 1658 bp in length
38576 38675: gap of 100 bp
38676 41309: contig of 2634 bp in length
41310 41409: gap of 100 bp
41410 43344: contig of 1935 bp in length
43345 43444: gap of 100 bp
43445 45853: contig of 2409 bp in length

```

```

* 45854 45953: gap of 100 bp
* 45954 48434: contig of 2481 bp in length
* 48435 48534: gap of 100 bp
* 48535 51026: contig of 2492 bp in length
* 51027 51126: gap of 100 bp
* 51127 54369: contig of 3243 bp in length
* 54370 54469: gap of 100 bp
* 54470 56661: contig of 2192 bp in length
* 56662 56761: gap of 100 bp
* 56762 59097: contig of 2336 bp in length
* 59098 59197: gap of 100 bp
* 59198 61551: contig of 2354 bp in length
* 61552 64257: contig of 2606 bp in length
* 64258 64357: gap of 100 bp
* 64358 67511: contig of 3154 bp in length
* 67512 67611: gap of 100 bp
* 67612 70329: contig of 2718 bp in length
* 70330 70429: gap of 100 bp
* 70430 73572: contig of 3143 bp in length
* 73573 73672: gap of 100 bp
* 73673 77203: contig of 3531 bp in length
* 77204 77303: gap of 100 bp
* 77304 80599: contig of 3296 bp in length
* 80600 80699: gap of 100 bp
* 80700 84046: contig of 3347 bp in length
* 84047 84146: gap of 100 bp
* 84147 88357: contig of 4211 bp in length
* 88358 88457: gap of 100 bp
* 88458 92454: contig of 3997 bp in length
* 92455 92554: gap of 100 bp
* 92555 96892: contig of 4338 bp in length
* 96893 96992: gap of 100 bp
* 96893 101334: contig of 4342 bp in length
* 101335 101434: gap of 100 bp
* 101435 105438: contig of 4004 bp in length
* 105439 105538: gap of 100 bp
* 105539 110413: contig of 4875 bp in length
* 110414 110513: gap of 100 bp
* 110514 115994: contig of 5481 bp in length
* 115995 116094: gap of 100 bp
* 116095 120447: contig of 4353 bp in length
* 120448 120547: gap of 100 bp
* 120548 124483: contig of 3936 bp in length
* 124484 124583: gap of 100 bp
* 124584 130249: contig of 5686 bp in length
* 130250 130349: gap of 100 bp
* 130350 135462: contig of 5113 bp in length
* 135463 135562: gap of 100 bp
* 135563 139847: contig of 4285 bp in length
* 139848 139947: gap of 100 bp
* 139948 146731: contig of 6784 bp in length
* 146732 146831: gap of 100 bp
* 146832 151469: contig of 4638 bp in length
* 151470 151569: gap of 100 bp
* 151570 159901: contig of 8332 bp in length
* 159902 160001: gap of 100 bp
* 160002 168308: contig of 8307 bp in length
* 168309 168408: gap of 100 bp
* 168409 176195: contig of 7787 bp in length
* 176196 176295: gap of 100 bp
* 176296 189303: contig of 13008 bp in length
* 189304 189403: gap of 100 bp
* 189404 202135: contig of 12732 bp in length
* 202136 202235: gap of 100 bp
* 202236 221365: contig of 19130 bp in length.

```

FEATURES  
 Location/Qualifiers  
 1..221365  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="RP11-462N18"  
 /clone\_lib="RPC1-11 Human Male BAC"  
 misc\_feature 1..1108

```
/note="assembly_fragment"
1209..2292
/note="assembly_fragment"
2393..3673
/note="assembly_fragment"
3774..4947
/note="assembly_fragment"
5048..6104
/note="assembly_fragment"
6205..7259
/note="assembly_fragment"
7360..8667

alignment_scores:
  Quality: 661.50      Length: 508
  Ratio: 3.410         Gaps: 8
  Percent Similarity: 38.189   Percent Identity: 34.252

alignment_block:
US-09-135-238b-2 x AC023534 ..
Align seg 1/1 to: AC023534 from: 1 to: 221365

2 AspArgTrpLeuTrp..... 6
|||||
33094 GACGGCTGGAGCTGGGAAGTAGATGGGGCTGGGATATAGTTAGAGATGGG 33143
7 ....ProLeuTyrPheLeuPro.....ValSerGlyAlaLeuArgI 19
|||||:|:|
33144 GCTCCCTTCGATTATATCTCTCTCCCTGCGAGTATCGGGGCCCTGAGGA 33193
19 leuLeuProGluValLysValGluGluGluLeuGlyLysSerValThrIle 35
|||||
33194 TCCTCCCAAGAGTAAAGGTAGAGGGGGAGCTGGGGGATCAGTTACCATC 33243
36 LysCysProLeuProGluMetHisValArgIleTyrLeuCysArgGluMe 52
|||||
33244 AAGTGGCCACTTCTCGAATGCATGTGAGGATATATCTGTGCGGGAGAT 33293
52 tAlaGlySerGlyThrCysGlyThrValValSerThrThrAsnPheIleL 69
|||||
33294 GCCTGGATCTGGAAACATGTGGTACCCTGGTGTATCCACCACCAACTTCATCA 33343
69 ysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProArgLysAsn 85
|||||
33344 AGGCAGATAACAAGGCCCGAGTTACTCTGAAGCAATACCCACGCAGAAAT 33393
86 LeuPheLeuValGluValThrGlnLeuThrGluSerAspSerGlyValTyr 102
|||||
33394 CTGTTCCTAGTGAGTAAACACAGCTGACAGAAAGTGACAGCGAGTCTA 33443
102 tAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLysValT 119
|||||
33444 TCCCTGGAGCGGCATGAACACAGACAGACGGGGGAAGACCCAGAAAGTCA 33493
119 hrLeuAsnValHis..... 123
|||||
33494 CCCTGAATGTGCACAGTGGTAGGTTCCTCCGCTGATGGAGGCTCAGACT 33543
123 ..... 123
33544 ACCCAGAAAATGTTTCACTTGGGAAGATAGGAACAGACATCTAACTGTGA 33593
123 ..... 123
33594 CAGAGTCTAGACTCTTACACTGTGGCCATATAAGCCACATATGATGTTGA 33643
123 ..... 123
33644 GATCCATATGAACCTCATATGATCAATTACCTCCATGTGTCACACAAATTTT 33693
123 ..... 123

33694 CTACCCAAACATGTATGAAAAGAACAGATAGGTGCTCTCTTTGTATTGTT 33743
123 ..... 123
33744 CATCTACCCAGGCGCATGAGGGGAGTAGAGGCTTGAGATTGAGGAGGCG 33793
123 ..... 123
33794 TCTGAGATGAGGAAATTTGGAAGCAAAAGAAATAACTATCCAAACGTAAT 33843
123 ..... 123
33844 TGTCCCATGAGAAATAAAGAGAAAGACACTAATTATCCAGCCACATTTAT 33893
123 ..... 123
33894 AGGCCAACTTATTACTCTGTGACCTGATCCTTTATTCTCTGGAACCCC 33943
123 ..... 123
33944 GCCTTTCTCCTTCTCTCCACTGAGTTCCCCACACATATGCAGCAGTAATT 33993
123 ..... 123
33994 CAACAGCCTAGAAATCCCAATGACCTCTGCTGGTTTTCGAGGGAAATGTGA 34043
123 ..... 123
34044 CTGGACTAGGGGAACCAAGTATTTCTCCAGGGGAAGCCCCCAACCCAGCTC 34093
123 ..... 123
34094 CTTCTCTGCTTCTTTTCAGTGGACCCTGGCCCTCATCCCTCAGTGTCTATCC 34143
123 ..... 123
34144 ACTTTGGCTGAAATAGTTCAGTTTACCCCAAGCCTCAGTCTCCCCCTA 34193
124 .....Ser..GluTyrGluProSerTr 130
:|:|:|
34194 GAGACCACCAACATATTCCTCTTCCTTCCATAGATAGATAGAGCCATCATG 34243
130 pGluGluGlnProMetProGluThrProLysTrpPheHisLeuProTyrL 147
|||||
34244 GGAAGAGCAGCAATGCTCGAGACTCCAAATGGTTTCATCTGCCCTATT 34293
147 euPheGlnMetProAlaTyrAlaSerSerSerLysPheValThrArg... 162
|||||
34294 TGTTCAGATGCTGCATATGCCAGTTCTTCCAAATTCGTAACCCAGAGGT 34343
163 ValThrThrProAlaGlnArgGlyLysValPro..... 173
|||:|:|:|
34344 CAGTTCACCATAGCAGGGTAGGAGGTTGGGTGAGATGCTGTGGCAATG 34393
174 .....ProValHisHisSerSerProThrT 182
|||||
34394 TCAAAATGGAGGGGGTCTATGTTCAATTCTAT..... 34427
182 hrGlnIleThrHisArgProArgValSerArgAlaSerSerValAlaGly 198
:|:|:|:|
34428 ..AGAGTGGCTGGCAGAGCTGAGCTGAGTCTTAGGTGTGTATTCCCTGCC 34475
199 AspLysProArgThrPheLeuProSerThrThrAlaSerLysIleSerAl 215
:|:|:|:|
34476 TTATAGAGAGGCACATACCTCCCTGAGATCTGGGTCATATCTCTCAAG 34525
215 alLeuGluGlyLeuLeuLysProGlnThrProSerTyrAsnHisHis..... 230
:|:|:|:|
34526 CCTGTCCGACAGTCAGTCCCTGGTCCGTAGCTGCAGAACTTACATATTA 34575
231 ..ThrArgLeuHisArgGlnArg 237
|||||
34576 CCACCAGGTGGCACCACCAAGCAA 34598
```

```
seq_name: gb_htg22:AC027719
seq_documentation_block:
LOCUS AC027719 58628 bp DNA HTG 01-APR-2000
DEFINITION Homo sapiens chromosome 8 clone RP11-258P9 map 8, LOW-PASS SEQUENCE
SAMPLING.
AC027719
VERSION AC027719.1 GI:7382531
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 58628)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 8, clone RP11-258P9
Unpublished
2 (bases 1 to 58628)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bida,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ginde,S., Goyette,M., Graham,L.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Melgrim,J., Meneus,L., Mihova,T., Miranda,C., Menga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,B., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (01-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L7449
Center clone name: 258_P_9
-----
* NOTE: This record contains 74 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1 671: contig of 671 bp in length
* 672 771: gap of 100 bp
* 772 1469: contig of 698 bp in length
* 1470 1569: gap of 100 bp
* 1570 2261: contig of 692 bp in length
* 2262 2361: gap of 100 bp
* 2362 3051: contig of 690 bp in length
*
3052 3151: gap of 100 bp
* 3152 3816: contig of 665 bp in length
* 3817 3916: gap of 100 bp
* 4597: contig of 681 bp in length
* 4598 4697: gap of 100 bp
* 4698 5395: contig of 698 bp in length
* 5396 5495: gap of 100 bp
* 5496 6185: contig of 690 bp in length
* 6186 6285: gap of 100 bp
* 6286 6969: contig of 684 bp in length
* 6970 7069: gap of 100 bp
* 7070 7780: contig of 711 bp in length
* 7781 7880: gap of 100 bp
* 7881 8590: contig of 710 bp in length
* 8591 8690: gap of 100 bp
* 8691 9372: contig of 682 bp in length
* 9373 9472: gap of 100 bp
* 9473 10162: contig of 690 bp in length
* 10163 10262: gap of 100 bp
* 10263 10944: contig of 682 bp in length
* 10945 11044: gap of 100 bp
* 11045 11747: contig of 703 bp in length
* 11748 11847: gap of 100 bp
* 11848 12547: contig of 700 bp in length
* 12548 12647: gap of 100 bp
* 12648 13336: contig of 689 bp in length
* 13337 13436: gap of 100 bp
* 13437 14103: contig of 667 bp in length
* 14104 14203: gap of 100 bp
* 14204 14872: contig of 669 bp in length
* 14873 14972: gap of 100 bp
* 14973 15682: contig of 710 bp in length
* 15683 15782: gap of 100 bp
* 15783 16491: contig of 709 bp in length
* 16492 16591: gap of 100 bp
* 16592 17290: contig of 699 bp in length
* 17291 17390: gap of 100 bp
* 17391 18088: contig of 698 bp in length
* 18089 18188: gap of 100 bp
* 18189 18883: contig of 695 bp in length
* 18884 18983: gap of 100 bp
* 18984 19671: contig of 688 bp in length
* 19672 19771: gap of 100 bp
* 19772 20413: contig of 642 bp in length
* 20414 20513: gap of 100 bp
* 20514 21196: contig of 683 bp in length
* 21197 21296: gap of 100 bp
* 21297 21992: contig of 696 bp in length
* 21993 22092: gap of 100 bp
* 22093 22804: contig of 712 bp in length
* 22805 22904: gap of 100 bp
* 22905 23604: contig of 700 bp in length
* 23605 23704: gap of 100 bp
* 23705 24402: contig of 698 bp in length
* 24403 24502: gap of 100 bp
* 24503 25191: contig of 689 bp in length
* 25192 25291: gap of 100 bp
* 25292 25974: contig of 683 bp in length
* 25975 26074: gap of 100 bp
* 26075 26769: contig of 695 bp in length
* 26770 26869: gap of 100 bp
* 26870 27572: contig of 703 bp in length
* 27573 27672: gap of 100 bp
* 27673 28373: contig of 701 bp in length
* 28374 28473: gap of 100 bp
* 28474 29165: contig of 692 bp in length
* 29166 29265: gap of 100 bp
* 29266 29956: contig of 691 bp in length
* 29957 30056: gap of 100 bp
* 30057 30741: contig of 685 bp in length
* 30742 30841: gap of 100 bp
* 30842 31519: contig of 678 bp in length
* 31520 31619: gap of 100 bp
```





ICQAGEPSADKKNVLDQVLAPELlyKDLRSVTPCEDLGREVANEAKYLCKMKNKE  
TCDVINTLGRDDPDEFGRILLITPKDDNGRFSVYLITGLRKEDAGHYOCGAHSSGLPQE  
GWPIQTQWLFNESTIPNRSVVKVGTGSAVATACYPNKESSSLKYWCWREGDNG  
HCPALVGTQAOVQREYEGRLALEDPQNGNYTVILNQLTTEDAGFTWCLNGSRWRT  
TLEQVAETREPNEVTPONATVIGLETFTVSPCKFYSEKQYKWCWKNKGCHIL  
PSHDEARQSSVCDQSOLVSMTLNPNVSKEDGWCYKQKQGTGETTAIYAYVEE  
TRGSSHYNPTDANARAKVLEEVEVDSSISEKENKAIPNPGPFANERETQNYRDOAQ  
ENRASGDGADGQSSSSSKVLFTPLVGLAVGAIAVWVAVRVHRKKNVDRMSIS  
SYRTDISMADFKNSRDLGNDNMGASPDQOTQVTEIGKDEIVTITTECTABPEESKKAKR  
SSKEADMAYSAFLLQSSSTAAQVHDGPQA"

mat\_peptide

misc\_feature

misc\_feature

misc\_feature

3'UTR

BASE COUNT

ORIGIN

..

Align seg 1/1 to: MMU06431 from: 1 to: 3095

10 PheLeuProValSerGlyAlaLeuArgIleLeuProGluValLysValG1 26

121 TTTTCAGGGCTCCACAAAACCCCATTTGGTCCCCAGAGGAGT 170

26 uGlyLeuLeuGlySerValThrIleLysCysProLeuProGlu... 41

171 TAGTATAGAGCGGACTCTGTTCCATCAGCTGCTACTACCCAGACACCT 220

42 .....MetHisValArgIleTyrLeuCysArgGluMetAlaGlySer 55

221 CTGTCAACCGGCACACCCGAAATACTGTTGCCGACAAGGAGCC...AGC 267

56 GlyThrCysGlyThrValValSerThrThrAspPheIleLysAlaGluTyr 72

268 GGATGTGACACACGCTCATCTCTTCAATGGCTACTCTCCAAAGGAGTA 317

72 rLysGlyArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuV 89

318 TTCAGGAGAGCCACCTCATCACTTCCAGAGACACACATTTGTGA 367

89 alGluValThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGly 105

368 TTAACATTGAGCAGCTCACCCAGGACACTGGGAGCTACAAGTGTGC 417

106 AlaGlyMetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnVa 122

418 CTGGGT...ACCAAGTAAACCGGCTGCTCTTCGATGTCAGCTGAGGT 464

122 lHisSerGluTyrGluProSerTyrGluGluGlnProMetProGluThrP 139

465 C.....AGCCAGGTTCCTGAGTTGC 484

139 roLysTrpPheHisLeuProTyrLeuPheGlnMetProLalaTyrAlaSer 155

485 CGAGTGACACCCAGCTC.....TACACAAAG 510

156 SerSerLysPheValThrArgValThrThrProAlaGlnArgGlyLysVa 172

511 GACATAGGCAGAAATGTGACCATTTGAATGCCCTTTTCAAAGGGAGAATGT 560

172 lPro 173

561 TCCC 564

seq\_name: gb\_om:AF091137

seq\_documentation\_block:

LOCUS AF091137 2848 bp mRNA MAM 01-JUL-1999

DEFINITION Trichosurus vulpecula polymeric-immunoglobulin receptor precursor

(PIGR) mRNA, complete cds.

ACCESSION AF091137

VERSION AF091137.1 GI:5305512

SOURCE Trichosurus vulpecula.

ORGANISM Trichosurus vulpecula

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Metatheria; Diprotodontia; Phalangeridae; Trichosurus.

REFERENCE 1 (bases 1 to 2848)

AUTHORS Adamski,F.M. and Demmer,J.

TITLE Cloning and characterization of pigR and J chain of the marsupial,

Trichosurus vulpecula (brush-tailed possum)

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 2848)

AUTHORS Adamski,F.M. and Demmer,J.

TITLE Direct Submission

JOURNAL Submitted (09-SEP-1998) Dairy Science, AgResearch Ruakura, East

Street, Hamilton, New Zealand

FEATURES Location/Qualifiers

source

1..2848

/organism="Trichosurus vulpecula"

/db\_xref="taxon:9337"

/tissue\_type="mammary gland"

1..2848

/gene="PIGR"

149..2350

/gene="PIGR"

/codon\_start=1

/product="polymeric-immunoglobulin receptor precursor"

/protein\_id="AAD41688.1"

/db\_xref="GI:5305513"

/translation="MAFLACLALLPVVSMKSPIFGKQVTVGGSGSVISQCFYPS

SYNRGRKRYFCLNQSCETIVSSNGFVSFRGRALTNFPGNNSFLIQISLEKE

DIGLYKCGLGTSNRGLSFDITLEVGEPLNPNTVEIVTEGKTVSINCPFOEQNTQD

RKFLACKDGESCALVIDSQVQPDYTGARLISIGTSRFEVTVISQIKRQDQNGYV

CGVGSDSTGIQKNVDLKLLEPEPELLYAGLGSGVTLNLCALGSTVASVPKFCQMRK

ETCDLVNSKGFNTNATHGRILFSTHTETGSPSIMTOVRKEDEGVYHCGAEDGOPS

EGTPRALQLFVSEETTPKSLPVKSGSGSVTITCDHPKKNNTLKYWKWEGSSH

CTKLVDLSGMDSEYEGRVALWDEPENGIITLNQTPDQDAGIYWCUSNGEHNKSS

VRIENDGQPLLIAPKTVTAQLGSLTISCHEPCKFYSEYKWKWSNQCGCTLPTQE

EGSSQAFVDCNQNSRNVSUTLNSVTRDHEGWYCGVKNQNGYGETIAVSVSVEEVSG

NAIOPTNAVLEDAVEPKVRGKEIEVPTDLGSTEBHSGSSVLTPLVPLALVTVGA

VALGIIKARWRFSRDSVSGSYRTDLSMSELENNPRQFGANENMDASVQETILGGEDE

IITQTEPPEIQQPKQVRKSSREEMATAITAFLLQSNMNEIPLDDPSNA"

BASE COUNT 769 a 690 c 722 g 667 t

ORIGIN

alignment\_scores:

Quality: 192.00 Length: 204

Ratio: 1.684 Gaps: 8

Percent Similarity: 55.882 Percent Identity: 30.392

alignment\_block:

US-09-135-238B-2 x AF091137 ..

Align seg 1/1 to: AF091137 from: 1 to: 2848

17 LeuArgIleLeuProGluVal..... 23

173 CTTCGCTCTCCCTCCCTGCTGTTCTTATGAGAGTCCCATATTTGCCCAA 222

24 .LysValGluGlyGluLeuGlyGlySerValThrIleLysCysProLeup 40

223 GCAAGTGACAGGAGTGGAAAGGAGGATCTGTCTCCATCCAGTCTTCTACC 272







```

KEYWORDS  JP 1998072495-A/1.
SOURCE      Homo sapiens.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1911)
AUTHORS     Miyabayashi,T. and Sakano,S.
TITLE       IMMUNITY-RELATED FACTOR
JOURNAL     Patent: JP 1998072495-A 17-MAR-1998;
            ASAHII CHEM IND CO LTD
COMMENT      OS Homo sapiens (human)
            PN JP 1998072495-A/1
            PD 17-MAR-1998
            PF 11-JUN-1997 JP 1997153218
            PR 13-JUN-1996 JP 96P 152362
            PI MIYABAYASHI TOMOYUKI, SAKANO SEIJI
            PC C07K14/47,C07H21/04,C07K16/18,C12N5/10,C12N15/09, PC
            C12P21/02//A61K38/00,
            PC C12P21/08, (C12N5/10,C12R1:91), (C12P21/02,C12R1:91), (C12P21/08,
            PC C12R1:91);
            CC strandedness: Double;
            CC topology: Linear;
            CC hypothetical: No;
            CC anti-sense: No;
            FH Key
            FH Location/Qualifiers
            FT source
            FT 1..1911
            FT /organism='Homo sapiens'
            FT /tissue_type='lymph node'
            FT 1..97
            FT 5'UTR
            FT 98..145
            FT /product='signal peptide of immunity related
            FT protein'
            FT 146..1693
            FT /product='immunity related protein' FT CDS
            FT 98..1696
            FT /product='immunity related protein' FT 3'UTR
            FT 1697..1911.
            FT Location/Qualifiers
FEATURES
source      1..1911
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
BASE COUNT  490 a 541 c 525 g 355 t
ORIGIN

alignment_scores:
  Quality: 187.50      Length: 575
  Ratio: 0.906        Gaps: 20
  Percent Similarity: 36.000  Percent Identity: 18.957

alignment_block:
US-09-135-238B-2 x E15470 ..
Align seg 1/1 to: E15470 from: 1 to: 1911

3 ArgTrpLeuTrpProLeuTyrPheLeuProValSerGlyAlaLeuArgI1 19
|||||
170 AGATGGCTGTGGAG...GGCTCTCTCCCTCCAGACCCATCTCCGGGC 216
19 LeuProGluValLys.....
:::
217 CATGGGAACATCAGGCCTTCCTCCGCCCTCTGCTGGGGGAGAGAGCT 266
25 .....ValGluGlyGlu 28
267 CCTTTGCAGCTCCAAATTCATTGAAGGGCTCAAGGCTGGTGTGTCAGGGGAG 316
29 LeuGlyGlySerValThrIleLysCysProLeuPro.....GL 41
|||||
317 CTGGAGGAGCTGTCCACCATCCAGTGCCCATTTATGCCCTCCATCTGTCAA 366
41 uMethisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrC 58

```

```

: ||| ||| ||| |||||
367 CAGGCACAGAGAGTACTGTGCGCTGGGGCCCCCAAGATGATCT 416
58 ysGlyThrValValSerThrThrAsnPheIleLysAlaGluTyrLysGly 74
|| |||||
417 GCAGACCATTTGTCTCCACCAACCATATATACCATCGCTATCGTGAC 466
75 ArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluVa 91
|||||
467 CGTGTGGCCCTCAGAGACTTCCACAGAGAGCTTGTGTGGTGAGGCT 516
91 IThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyM 108
:::
517 GTCCCAACTGTCCCGGATGACATCGGATGCTACCTCTCGCGCATGGAA 566
108 etAsnThrAspArg.....
:::
567 GTGAACAACATGCTGTTCTTAAGCATGAATCTGACCATCTCTGCAGGT 616
112 .....
617 CCCGCCAGACCTCCCCACAGCCACTCCAGCTGCTGGGGAGCTCACCAT 666
113 .....GlyL 114
667 GAGATCCTATGAAACAGCGCTCTCCAGTGGCCAAACAGATGGACCCAGAA 716
114 ysThrGln.....
717 CCACCCAGACCTTAGGACAGGGGACAGCATGGACACAGTGTCTTCCACT 766
117 .....LysValThrLeuAsnValHisSerGluTyrGluPro... 128
|||
767 CCAGGAACAGCAAGACTACAGCTTCAGCTGAGGAAGACGAACCCAGG 816
129 .....SerTirpGluGlu..... 132
|||||
817 AGCAACACAGGCAGCGCTCCAGGACAGGAGCTGGGAGAGGGTCTG 866
133 .....GlnProMetProGluThrProLysTrpPheHisLeuPro 145
|||||
867 TCAAGACACCTGCTCCGATTTCCAGAGAGTCCA..... 898
146 TyrLeuPheGlnMetProAlaTyrAlaSerSerLysPheValThrAr 162
|||||
899 .....CCTTCAAAGACAGAAAGCATGTCCAA 924
162 gValThrThrProAlaGlnArgGlyLysValProProValHisHisSerS 179
:::
925 TACAACAGAGGTGTTGGAGGGC..... 949
179 erProThrThrGlnIleThrHisArgProArgValSerArg..... 192
:::
950 ..ACCAGAAAGCTCGGTGACAAACAGGGCTAGAGCCAGCAAGGACAGGAG 997
193 ..AlaSerSerValAlaGlyAspLysProArg..... 202
:::
998 GAGATGACAACTACCAAGGCTGATGGCCAGGAGGACATAGAGGGGT 1047
202 .....
1048 CAGGATAGCTCTTGATGAGCCCAAAAGGTCTAGAACCATTTGGCCAC 1097
203 .....ThrPheLeuProSerThrThr 209
|||
1098 CAGCTCTGTGTCAGAAACTTTGGCGTGGGAAATCCTCCCAACAAGCAAG 1147
210 AlaSerLysIleSerAlaLeuGluGlyLeuLysProGlnThrProSe 226
:::
1148 CCAGTTTCTTAAGCAACAATCTCAGGGTTCCATTGGAGAAACAACCTCAGC 1197
226 rTyrAsnHisHisThrArg..... 232
:::
|||

```

```
1198 TGCAGCATGTGGACCTTGGAACTCCAGCTGCAGATGTGTGATCTTGG 1247
233 .....LeuHisArgGlnArgAlaLeuAspTyr 241
1248 GAATCCAGCTGCAGATGTGTGGACCATGTGGAGCAGCATCTGGGAA 1297
242 GlycerGlnSerGly..... 246
1298 GGAAGCGCTGCAGGGACCTAGATGTGCCACTGGACAGAGTCCCCA 1347
246 ..... 246
1348 AGCAACACTGACCCAGACCCCGGCAGTAGGACCTGGGACCCCTGGCA 1397
247 .....ArgGluGlyGlnGlyPheHis 253
1398 AGAGTCCTCCGTGAAGCGTACTTTTCCAGAAAGATGAAGCAGCTCTCG 1447
254 IleLeuIlePro.....ThrIleLeuGlyLeuPheLeuAlaLeuLe 268
1448 ACCCTGGCTCTGTCTCTACCATGTGGCCCTGTTATGCTTATGGCTCT 1497
268 uGlyLeuValVallYsArgAlaValIleuArgGlyAlaLeuSerArgA 285
1498 GGTCTCTATTG.....CAAAGGAAGCTCTGGAGAAGGA 1529
285 rgAlaArgArgLeuAlaValArgMetArgAlaLeuGluSerSerGln... 300
1530 GGACCTCTCAGAGGAGCAAGGGTCACCTTAATTCAGATGACATTTT 1579
301 .....ArgProArgGlySerProArgProArgSerGlnAsnAl 314
1580 CTGGAAGTGAACCCCAAGCAGACGAGCTGCCCATGTGGAAGAAGAT 1629
314 eTyr.....SerAlaCysProArgArgAlaArgGlyAlaAspAlaAlaG 329
1630 GTCCAGGATGACTCTCTCTCT.....GCTGGGGCCAGCC 1664
329 lyThrGlyAlaProValProGlyPro.....Gly 339
1665 TGACTGCCCCAGAGAGAAATCCAGGACCTGAGGAGACAGATGAAT 1714
340 AlaProLeuPro.....ProAl 345
1715 GTCAGTATACATGGGAGAGACCAAGATCAAGGCTTCAGGACCCCA 1764
345 aProLeuGlnValSerGluSerPro 353
1765 GCCTCTTTCCATCATCTCTCTCCA 1789

seq_name: gb_ro:RNIGR
seq_documentation_block:
LOCUS RNIGR 3269 bp mRNA ROD 22-MAR-1995
DEFINITION Rat mRNA for polymeric immunoglobulin receptor.
ACCESSION X15741
VERSION X15741.1 GI:56464
KEYWORDS immunoglobulin receptor; polymeric immunoglobulin receptor;
transmembrane protein.
SOURCE
Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE
1 (bases 1 to 3269)
AUTHORS Banting G.
TITLE Direct Submission
JOURNAL Submitted (05-JUL-1989) Banting G., EMBL, Meyerhofstr 1, 6900
Heidelberg
REFERENCE
2 (bases 1 to 3269)
AUTHORS Banting G., Brake B., Braghetta P., Luzio J.P. and Stanley K.K.
JOURNAL Unpublished
FEATURES
source 1. 3269
```

```
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/tissue_type="liver"
/clone_lib="cDNA, TR11"
/clone="GORFL8,43,4339"
74. .127
/product="rat polymeric immunoglobulin receptor signal
peptide (AA -18 to -1)"
74. .2383
/note="rat polymeric immunoglobulin receptor (AA -18 to
751)"
/codon_start=1
/protein_id="CAA33758.1"
/db_xref="GI:56465"
/db_xref="SWISS-PROT:P15083"
/translation="MRLSLFALLVTFSGVSTQSPIFGPQDVSSIEGNSVSICTYPD
TSVNRTRYMCWGQAGYCATLISSNGYLSKSGRASLINFNPENSTFVINIAHLTO
EDTSYKCGLTNRGLFFDVSQVPEPNDTHVTTKDGRVTTECFRKEGNAH
SKSLCKRGEACVWIDSTEYVDSYKDRALLPMKTSRDIIFYVNIHSLIPSDAGLY
VCQAGGSPADKNNADLOVLEPEPELLYKDLSSVTECDLGREVANDAKYLCKRNKE
TCDVINTLGRDPAFEGRIILLTPRDDNGRFSLITLKRKEDAGHYQCCAGHSSGLPQE
GWPVQALQFVNEESTIPNSRVTYKGVTSVAIVCPYNPKESSLKYWCHEADENG
RCPLVGTQALVQEGYGRALAFDQPGSGATVILNQLTTQDSGFYWCILTQDSRWRT
TIELOVAEATKKPDLVTPONATVETFTISCHYPCKEYSQRYCKWNSDGSCHIL
PSHDEGAROSVSCDOSQIVSMTLPNVKDEGWYCWKGQVYGETTATYVAVEE
RTRGSPHINTDANARAKDAPEEAMESSVREDENKALDPRLFADERIONAGDQAO
ENRASNAGSAGGQSGSKVLFSTPLVGLVLAGVAVVAVRHRKRVDMRMSISSY
RTDISMGDFRNSRDJGGNDMGATPDQTETVLEGRDEIETITTECTTEPESKAKRRS
KEADWYSAFLFOSSTIAAOVHDGPOEA"
128. .2380
/product="rat polymeric immunoglobulin receptor mature
peptide (AA 1 to 751)"
BASE COUNT 890 a 831 c 850 g 598 t
ORIGIN

alignment_scores:
Quality: 185.50 Length: 164
Ratio: 1.855 Gaps: 6
Percent Similarity: 60.976 Percent Identity: 30.488

alignment_block:
US-09-135-238B-2 x RNIGR ..
Align seg 1/1 to: RNIGR from: 1 to: 3269

30 GlyGlySerValThrIleLysCysProLeuProGlu.....Me 42
170 GGTAACTCGTCTCCATCAGCTGCTACTACCCAGACACCTCTGTCAACCG 219

42 ThisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 59
220 GCACACCCGAAATACTGGTCCCAAGAGGCC...AACGGCTACTGGC 266

59 lyThrValValSerThrThrAsnPheIleLysAlaGluTyrLysGlyArg 75
267 CAACCCCTCATCTCTCAATGGCTACCTCTCGAGGAGTATTCAGGACAGA 316

76 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValth 92
317 GCCAGCCTCATCAACTCCAGAGATAGCACATTTGTGATTACATTCG 366

92 rGlnLeuThrGluSerSerGlyValTyrAlaCysGlyAlaGlyMetA 109
367 ACATCTCACCAGGAGGACACTGGAGCTACAAGTGTGCTGGGT...A 413

109 snThrAspArgGlyLysThrGlnLysValThrIleuAsnValHisSerGlu 125
414 CCACATACCAGGACCTGTTTTCGATGTCAGCCTGGAGGTC..... 454

126 TyrGluProSerTrpGluGlnProMetProGluThrProLysTrpPh 142
455 .....AGCCAGGTTCTCTGAGTCCCAATGACAC 483
```







LOCUS S43435S03 345 bp DNA PRI 08-MAY-1993  
 DEFINITION transmembrane secretory component [human, leukocytes, Genomic, 345 nt, segment 3 of 11].  
 ACCESSION S43441  
 VERSION S43441.1 GI:255089  
 KEYWORDS 3 of 11  
 SEGMENT human leukocytes.  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 345)  
 AUTHORS Krajci,P., Kvale,D., Tasken,K. and Brandtzaeg,P.  
 TITLE Molecular cloning and exon-intron mapping of the gene encoding human transmembrane secretory component (the poly-Ig receptor)  
 JOURNAL Eur J. Immunol. 22 (9), 2309-2315 (1992)  
 MEDLINE 92387236  
 REMARK GenBank staff at the National Library of Medicine created this entry [NCBI gibbsq 113232] from the original journal article. This sequence comes from Fig. 2.  
 FEATURES  
 source Location/Qualifiers  
 1..345 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 BASE COUNT 78 a 102 c 97 g 68 t  
 ORIGIN  
 alignment\_scores  
 Quality: 182.00 Length: 115  
 Ratio: 2.247 Gaps: 5  
 Percent Similarity: 70.435 Percent Identity: 37.391  
 alignment\_block  
 US-09-135-238b-2 x S43435S03 ..  
 Align seg 1/1 to: S43435S03 from: 1 to: 345  
 12 ProValSerGlyAlaLeuArgIleLeuProGluValLysValGluGlycl 28  
 |||||: |||  
 18 CCATATTGGT.....CCCGAG...GAGGTGAATAGTGT 49  
 28 uLeuGlySerValThrIleLysCysProLeuPro.....G 41  
 |||||: |||||: |||  
 50 GGAAGCTAACTCAGTGTCCATCAGTGTCTACTACCCACCCCTCTGTCA 99  
 41 luMetHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThr 57  
 :: ||||: ||| ||| ||||: ||| |||  
 100 ACCGCGCACCCCGGAAGTACTGGTCCGCGCAGGAGCTAGAGGTGGC... 146  
 58 CysGlyThrValValSerThrThrAsnPheIleLysAlaGluTyrLyscl 74  
 ||| ||||: ||||: |||: ||||: ||||: ||||  
 147 TGCATAACCTCATCTCTCGAGGGCTACGCTCCAGCAATATGCGAG 196  
 74 YargValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluV 91  
 ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||  
 197 CAGGCTAACTCACCACCTCCCGAGAGCGCACATTGTGGTGAACA 246  
 91 alThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGly 107  
 ::|||: ||||: ||||: ||||: ||||: ||||: ||||: ||||  
 247 TTGCCAGCTAGCCAGGATGACTCCGGCGCTACAAGTGTGGCTGGGC 296  
 108 MetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnVal 122  
 ::|||: ||||: ||||: ||||: ||||: ||||: ||||: ||||  
 297 ATCAATAGC...CGAGGCTGTCTTTGATGTACGCTGGAGGTC 338  
 seq\_name: gb\_pat:A52094  
 seq\_documentation\_block:  
 LOCUS A52094 1936 bp DNA PAT 11-MAR-1997  
 DEFINITION Sequence 4 from Patent WO9618734.  
 ACCESSION A52094  
 VERSION A52094.1 GI:2304706

KEYWORDS unidentified.  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 REFERENCE 1 (bases 1 to 1936)  
 AUTHORS Van,O.J., Asselbergs,F.A., Leist,C. and Gruetter,M.  
 TITLE PRODUCTION OF RECOMBINANT SECRETORY COMPONENT  
 JOURNAL Patent: WO 9618734-A 4 20-JUN-1996;  
 CIBA GEIGY AG (CH)  
 COMMENT Other publication AU 4302796 960703.  
 FEATURES  
 source Location/Qualifiers  
 1..1936 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 16..1920  
 mat\_peptide  
 BASE COUNT 451 a 521 c 583 g 381 t  
 ORIGIN  
 alignment\_scores  
 Quality: 182.00 Length: 115  
 Ratio: 2.247 Gaps: 5  
 Percent Similarity: 70.435 Percent Identity: 37.391  
 alignment\_block  
 US-09-135-238b-2 x A52094 ..  
 Align seg 1/1 to: A52094 from: 1 to: 1936  
 12 ProValSerGlyAlaLeuArgIleLeuProGluValLysValGluGlycl 28  
 |||||: |||  
 76 CCATATTGGT.....CCCGAG...GAGGTGAATAGTGT 107  
 28 uLeuGlySerValThrIleLysCysProLeuPro.....G 41  
 |||||: |||||: |||  
 108 GGAAGTAACTCAGTGTCCATCAGTGTCTACTACCCACCCCTCTGTCA 157  
 41 luMetHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThr 57  
 :: ||||: ||| ||| ||||: ||| |||  
 158 ACCGCGCACCCCGGAAGTACTGGTCCGCGCAGGAGCTAGAGGTGGC... 204  
 58 CysGlyThrValValSerThrThrAsnPheIleLysAlaGluTyrLyscl 74  
 ||| ||||: ||||: |||: ||||: ||||: ||||  
 205 TGCATAACCTCATCTCTCGAGGGCTACGCTCCAGCAATATGCGAG 254  
 74 YargValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluV 91  
 ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||  
 255 CAGGCTAACTCACCACCTCCCGAGAGCGCACATTCTGGTGAACA 304  
 91 alThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGly 107  
 ::|||: ||||: ||||: ||||: ||||: ||||: ||||: ||||  
 305 TTGCCAGCTAGCCAGGATGACTCCGGCGCTACAAGTGTGGCTGGGC 354  
 108 MetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnVal 122  
 ::|||: ||||: ||||: ||||: ||||: ||||: ||||: ||||  
 355 ATCAATAGC...CGAGGCTGTCTTTGATGTACGCTGGAGGTC 396

-----

OM of: US-09-135-238B-2\_COPY\_18\_253 to: GenEmbl.\* out\_format : pfs  
Date: Sep 12, 2000 6:53 AM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlp  
-O=/cgn2\_1/USPTO.spool/US09135238/runat\_24082000\_091323\_17880/app\_query.fasta\_1.1282  
-DB=GenEmbl -OFMT=fastap -SUFFIX=p2n.rge -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-OGAPO=4.000 -OGAPEXT=0.000 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPO=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62  
-TRANS=human40 cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM=ext -MINLEN=0 -MAXLEN=1000000  
-USER=US09135238\_@CGNL\_1\_5120 -NCPU=6 -ICPU=3 -LONGLOG -NO\_XLPXY  
-WAIT -THREADS=1

Search information block:  
Query: US-09-135-238B-2\_COPY\_18\_253  
Query length: 236  
Database: GenEmbl.\*  
Database sequences: 972840  
Database length: 892348106  
Search time (sec): 2784.300000

score_list:	Strd Orig	ZScore	EScore	Len	Documentation
gb_pr3:AF057557	+ 1254.00	1700.48	1.3e-86	1339	! AF057557 Homo sapiens anti-Fas
gb_h9g11:AC023534	+ 629.50	806.66	7.8e-37	221365	! AC023534 Homo sapiens clone
gb_h9g22:AC027719	- 252.00	304.59	7.2e-09	58628	! AC027719 Homo sapiens chromos
gb_ro:MM06431	+ 197.00	255.23	4.0e-06	3095	! U06431 Mus musculus B6/CBAF1J
gb_om:AF091137	+ 189.00	245.06	1.5e-05	2417	! AF091137 Trichosurus vulpecula
gb_ro:MMPIGR03	+ 188.00	260.34	2.1e-06	418	! U83428 Mus musculus polymeric i
gb_ro:MMPIGR3	+ 188.00	258.68	2.6e-06	505	! Y16525 Mus musculus p1gr gene,
gb_pat:EI5253	+ 188.00	223.50	0.0002	29392	! EI5253 Mouse pig receptor gen
gb_ro:AB001489	+ 188.00	223.50	0.0002	29392	! AB001489 Mus musculus DNA for
gb_om:RNIGR	+ 185.50	239.11	3.2e-05	3269	! X15741 Rat mRNA for polymeric
gb_om:AC032195	+ 181.00	235.44	5.1e-05	2461	! AB032195 Sus scrofa mRNA for p
gb_pr2:S43435S03	+ 180.50	251.77	6.3e-06	345	! S43441 transmembrane secretory
gb_pat:A52094	+ 180.50	236.84	4.3e-05	1936	! A52094 Sequence 4 from Patent
gb_pat:A52093	+ 180.50	236.43	4.5e-05	2031	! A52093 Sequence 3 from Patent
gb_pat:A52091	+ 180.50	234.96	5.4e-05	2405	! A52091 Sequence 1 from Patent
gb_pr2:S62403	+ 180.50	233.35	6.7e-05	2898	! S62403 transmembrane secretory
gb_pr1:HSPIR	+ 180.50	233.29	6.7e-05	2919	! X73079 Homo sapiens encoding p
gb_om:BTPIGR1	+ 179.50	230.04	0.0001	3630	! X81371 B.taurus mRNA for bovin
gb_om:BOVPIGR	+ 179.00	229.61	0.0001	3527	! L04797 Bos taurus polymeric im
gb_pat:EI5470	+ 175.50	230.15	0.0001	1911	! EI5470 cDNA encoding immunity
gb_om:OCIGRECR	+ 166.50	212.63	0.0010	3517	! X00412 Rabbit mRNA for poly-im
gb_h9g11:AC023534	- 166.00	176.09	0.1035	221365	! AC023534 Homo sapiens clone
gb_ro:MMPIGR5	+ 146.50	203.10	0.0032	456	! Y16527 Mus musculus p1gr gene,
gb_pr1:HUMIGRPOLY	+ 143.00	183.52	0.0399	2526	! M24559 Human poly-Ig receptor
gb_ro:MMPIGR05	+ 137.00	189.72	0.0180	481	! U83430 Mus musculus polymeric i
gb_pr2:S43435S05	+ 132.50	186.78	0.0263	333	! S43443 transmembrane secretory
gb_pr2:S43435S04	+ 130.00	177.49	0.0864	657	! S43442 transmembrane secretory
gb_pr2:HS2438323	+ 124.00	165.97	0.3791	969	! AJ238323 Homo sapiens mRNA for
gb_pr2:HS2410101	+ 124.00	165.96	0.3795	970	! AJ010101 Homo sapiens mRNA for
gb_pr2:HS2424864	+ 124.00	160.34	0.7805	1857	! AJ224864 Homo sapiens mRNA for
gb_h9g29:AC060797	- 121.00	126.24	61.86	59506	! AC060797 Homo sapiens chromos
gb_h9g15:AC016888	- 121.00	116.89	205.24	175260	! AC016888 Homo sapiens chromid
gb_pr3:AF043724	+ 117.50	153.70	1.83	1440	! AF043724 Homo sapiens hepatic
gb_om:RABTCBXB	+ 114.50	150.17	2.88	1351	! M13895 Rabbit T-cell receptor
gb_ro:AF035963	+ 111.50	140.62	9.78	2539	! AF035963 Rattus norvegicus kid
gb_pr1:HSNCAMX1	+ 111.00	123.85	84.04	16288	! Z29373 H.sapiens gene for neu
gb_bat:AF000062	+ 110.50	99.47	1.9e+03	251700	! AP000062 Aeropyrum pernix ge
gb_h9g22:AC027719	+ 109.50	110.73	452.59	58628	! AC027719 Homo sapiens chromos
gb_h9g5:AC013305	- 109.50	101.22	1.5e+03	175799	! AC013305 Homo sapiens clone
gb_pr2:HUMTCBYZ	+ 109.00	144.67	5.82	1074	! M15564 Human T-cell receptor r

gb\_pr2:HUMTCRBAAL + 108.50 145.18 5.45 936 ! L34734 Homo sapiens T-cell r  
gb\_h9g28:AC021643 - 108.00 96.02 3.0e+03 253327 ! AC021643 Mus musculus chr  
gb\_ro:MMPIGR04 + 107.50 145.84 5.01 741 ! U83429 Mus musculus polymeri  
gb\_ro:MMPIGR4 + 107.50 145.22 5.42 796 ! Y16526 Mus musculus p1gr gen  
gb\_ov:D85084 + 107.50 130.43 36.17 4398 ! D85084 Cynops pyrrhogaster  
  
seq\_name: gb\_pr3:AF057557  
  
seq documentation block:  
LOCUS AF057557 1339 bp mRNA PRI 04-JUN-1998  
DEFINITION Homo sapiens anti-Fas-induced apoptosis (TOSO) mRNA, complete cds.  
ACCESSION AF057557  
VERSION AF057557.1 GI:3169292  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1339)  
AUTHORS Hitoshi,Y., Lorens,J., Kitada,S.I., Fisher,J., LaBarge,M.,  
Ring,H.Z., Francke,U., Reed,J.C., Kinoshita,S. and Nolan,G.P.  
TITLE Toso, a cell surface, specific regulator of Fas-induced apoptosis  
in T cells  
JOURNAL Immunity 8 (4), 461-471 (1998)  
MEDLINE 98246048  
REFERENCE 2 (bases 1 to 1339)  
AUTHORS LaBarge,M. and Hitoshi,Y.  
TITLE Direct Submission  
JOURNAL Submitted (03-APR-1998) Molecular Pharmacology, Stanford  
University, 300 Pasteur Dr., Palo Alto, CA 94304, USA  
  
FEATURES  
Location/Qualifiers  
1..1339  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="lq31-q32"  
/cell\_line="Jurkat"  
/cell\_type="T cell"  
/note="TOSO.2 cDNA; cloned by PCR using primers designed  
from the original TOSO.1 cDNA sequence"  
1..1339  
/gene="TOSO"  
20..1192  
/gene="TOSO"  
/note="transmembrane protein; Ig superfamily member"  
/codon\_start=1  
/product="anti-Fas-induced apoptosis"  
/protein\_id="AAC18830.1"  
/db\_xref="GI:3169293"  
/translation="MDFWLPFLFVSGALRILPEVKVEGELGGSVTKCPLEPMHV  
RILYLCRMAGSGTCGTVTNFIKAEYKGRVTLKQYPRKNLFLVEVTLQTESDGVY  
ACGAGMTRDRGKTQVTLNVHSEYEWEEQPMETPKWHLPLYFQMPAYASSKFEV  
TRVTPAPGRKVPVYHSSPTQITHRPRVSRASSVAGDKPRTFLPFTASKISALEG  
LKPTPTPSYNNHTRLHRQALDYSGSGREGQPHILPTILGLIFLALLGLVWKRVA  
ERRKALSRARLEAVRMALRESSQRPGRSPRSONNIYSACPRRACDAAGTGEAP  
VPGCAPLPPAPLQVSESPWLHAPSLKTSCEVSLYHQHPAMMEDSDDDYINWPA"  
127  
variation  
/gene="TOSO"  
/note="occurs in TOSO.1 cDNA"  
/replace="a"  
1015  
variation  
/gene="TOSO"  
/note="occurs in TOSO.1 cDNA"  
/replace="g"  
BASE COUNT 309 a 432 c 340 g 258 t  
ORIGIN  
  
alignment\_scores:  
Quality: 1254.00 Length: 236  
Ratio: 5.314 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-135-238B-2\_Copy\_18\_253 x AF057557

Align seg 1/1 to: AF057557 from: 1 to: 1339

```
1  ArgIleLeuProGluValLysValGluGlyLeuGlyCysValTh 17
|||||
71  AGGATCTCCAGAGTAAGGTAGAGGGGAGCTGGCGGATCATGTTAC 120
|||||
17  rIleLysCysProLeuProGluMetHisValArgIleTyrLeuCysArg 34
|||||
121  CATCAGTGGCCACTTCTCGAATGCATGTGAGATATATCTGTCCGGG 170
|||||
34  luMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsnPhe 50
|||||
171  AGATGGCTGGATCTGGAACATGTGTACCGTGGTATCCACCACTTC 220
|||||
51  iLeuAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProArgLy 67
|||||
221  ATCAGGCGAGATACAGGGCGGAGTCTCTGAAGCAATACCCAGCAA 270
|||||
67  sAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerGlyV 84
|||||
271  GAATCTGTCTCTAGTGGAGGTATACACAGCTGACAGAAAGTGACGCGGAG 320
|||||
84  alTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLys 100
|||||
321  TCTATCGCTGGCGGCGGCGATGACACAGACCGGGGAGAGACCCAGAAA 370
|||||
101  ValThrLeuAsnValHisSerGluTyrGluProSerTrpGluGluInPr 117
|||||
371  GTCCACCTGAATGTCCACAGTGAATACAGGCCATCATGGGAGAGCAGCC 420
|||||
117  oMetProGluThrProLysTrpPheHisLeuProTyrLeuPheGlnMetP 134
|||||
421  AATGCTGTGAGACTCCAAATGGTTTCATCTGCCCTATTGTTCCAGATGC 470
|||||
134  roAlaTyrAlaSerSerSerLysPheValThrArgValThrThrProAla 150
|||||
471  CTCATATGCCAGTCTCTCCAAATTCGTAAACGAGATTACACACAGCT 520
|||||
151  GlnArgGlyLysValProValHisHisSerSerProThrThrGlnIle 167
|||||
521  CAAAGGGGCAAGTCCCTCCAGTTCACCACTCTCCCTCCACCAACCAAT 570
|||||
167  eThrHisArgProArgValSerArgAlaSerSerValAlaGlyAspLysP 184
|||||
571  CACCCACCGCCCTCGAGTGTCCAGAGCATCTCAGTAGCAGGTGCAAGC 620
|||||
184  roArgThrPheLeuProSerThrThrAlaSerLysIleSerAlaLeuGlu 200
|||||
621  CCGGAACCTTCTGCCATCCACTACAGCTCAAAAATCTCAGCTCTGGAG 670
|||||
201  GlyLeuLeuLysProGlnThrProSerTyrAsnHisThrArgLeuHi 217
|||||
671  GGCTGCTCAAGCCCGCCAGCGCCAGCTACAAACCCACACACAGGCTGCA 720
|||||
217  sArgGlnArgAlaLeuAspTyrGlySerGlnSerGlyArgGluGlyGlnG 234
|||||
721  CAGGACAGAGCACTGGCACTATGGCTACAGTCTGGGAGGAGGCCAG 770
|||||
234  lyPheHis 236
|||||
771  GATTTTAC 778
```

seq\_name: gb\_htg11:AC023534

seq\_documentation\_block:

LOCUS AC023534 221365 bp DNA HTG 02-MAR-2000  
DEFINITION Homo sapiens clone RP11-462N18, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 61  
unordered pieces.  
ACCESSION AC023534  
VERSION AC023534.2 GI:7143452

KEYWORDS  
SOURCE  
ORGANISM

HTG: HTGS\_PHASE1.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 221365)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Homo sapiens chromosome, clone RP11-462N18

Unpublished

2 (bases 1 to 221365)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,

Bouckgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,

Choepeil,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,

DeArelano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,

Fenster,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,

Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,

Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,

Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,

Klein,J., Landers,T., Largocque,K., Lehoczy,J., Levine,R.,

Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,

McEwan,P., McGurk,A., McKernan,K., McPheeters,K., Meldrum,J.,

Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J., Naylor,J.,

Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivari,T.M.,

Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,

Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,

Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,

Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,

Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,

Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and

Zody,M.

Direct Submission

Submitted (15-FEB-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Mar 2, 2000 this sequence version replaced gi:6978210.

All repeats were identified using RepeatMasker:

Smit, A.F.A. &amp; Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L5244

Center clone name: 462\_N\_18

-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 61 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 1108: contig of 1108 bp in length  
\* 1109 1208: gap of 100 bp  
\* 1209 2292: contig of 1084 bp in length  
\* 2293 2392: gap of 100 bp  
\* 2393 3673: contig of 1281 bp in length  
\* 3674 3773: gap of 100 bp  
\* 3774 4947: contig of 1174 bp in length  
\* 4948 5047: gap of 100 bp  
\* 5048 6104: contig of 1057 bp in length  
\* 6105 6204: gap of 100 bp  
\* 6205 7259: contig of 1055 bp in length  
\* 7260 7359: gap of 100 bp  
\* 7360 8667: contig of 1308 bp in length  
\* 8668 8767: gap of 100 bp  
\* 8768 10566: contig of 1799 bp in length  
\* 10567 10666: gap of 100 bp  
\* 10667 12173: contig of 1507 bp in length  
\* 12174 12273: gap of 100 bp

```
* 12274 13643: contig of 1370 bp in length
* 13644 13743: gap of 100 bp
* 13744 15215: contig of 1472 bp in length
* 15216 15315: gap of 100 bp
* 15316 16609: contig of 1294 bp in length
* 16610 16709: gap of 100 bp
* 16710 18085: contig of 1376 bp in length
* 18086 18185: gap of 100 bp
* 18186 19668: contig of 1483 bp in length
* 19669 19768: gap of 100 bp
* 19769 22083: contig of 2315 bp in length
* 22084 22183: gap of 100 bp
* 22184 24052: contig of 1869 bp in length
* 24053 24152: gap of 100 bp
* 24153 25902: contig of 1750 bp in length
* 25903 26002: gap of 100 bp
* 26003 27639: contig of 1637 bp in length
* 27640 27739: gap of 100 bp
* 27740 29377: contig of 1638 bp in length
* 29378 29477: gap of 100 bp
* 29478 30910: contig of 1433 bp in length
* 30911 31010: gap of 100 bp
* 31011 31225: contig of 215 bp in length
* 31226 31325: gap of 100 bp
* 31326 32890: contig of 1565 bp in length
* 32891 32990: gap of 100 bp
* 32991 34812: contig of 1822 bp in length
* 34813 34912: gap of 100 bp
* 34913 36817: contig of 1905 bp in length
* 36818 36917: gap of 100 bp
* 36918 38575: contig of 1658 bp in length
* 38576 38675: gap of 100 bp
* 38676 41309: contig of 2634 bp in length
* 41310 41409: gap of 100 bp
* 41410 43344: contig of 1935 bp in length
* 43345 43444: gap of 100 bp
* 43445 45853: contig of 2409 bp in length
* 45854 45953: gap of 100 bp
* 45954 48434: contig of 2481 bp in length
* 48435 48534: gap of 100 bp
* 48535 51026: contig of 2492 bp in length
* 51027 51126: gap of 100 bp
* 51127 54369: contig of 3243 bp in length
* 54370 54469: gap of 100 bp
* 54470 56681: contig of 2192 bp in length
* 56682 56781: gap of 100 bp
* 56782 59097: contig of 2336 bp in length
* 59098 59197: gap of 100 bp
* 59198 61551: contig of 2354 bp in length
* 61552 61651: gap of 100 bp
* 61652 64257: contig of 2606 bp in length
* 64258 64357: gap of 100 bp
* 64358 67511: contig of 3154 bp in length
* 67512 67611: gap of 100 bp
* 67612 70329: contig of 2718 bp in length
* 70330 70429: gap of 100 bp
* 70430 73572: contig of 3143 bp in length
* 73573 73672: gap of 100 bp
* 73673 77203: contig of 3531 bp in length
* 77204 77303: gap of 100 bp
* 77304 80599: contig of 3296 bp in length
* 80600 80699: gap of 100 bp
* 80700 84045: contig of 3347 bp in length
* 84047 84146: gap of 100 bp
* 84147 88357: contig of 4211 bp in length
* 88358 88457: gap of 100 bp
* 88458 92454: contig of 3997 bp in length
* 92455 92554: gap of 100 bp
* 92555 96893: contig of 4338 bp in length
* 96893 96992: gap of 100 bp
* 96993 101334: contig of 4342 bp in length
* 101335 101434: gap of 100 bp
* 101435 105438: contig of 4004 bp in length
```

```
* 105439 105538: gap of 100 bp
* 105539 110413: contig of 4875 bp in length
* 110414 110513: gap of 100 bp
* 110514 115994: contig of 5481 bp in length
* 115995 116094: gap of 100 bp
* 116095 120447: contig of 4353 bp in length
* 120448 120547: gap of 100 bp
* 120548 124483: contig of 3936 bp in length
* 124484 124583: gap of 100 bp
* 124584 130249: contig of 5666 bp in length
* 130250 130349: gap of 100 bp
* 130350 135462: contig of 5113 bp in length
* 135463 135562: gap of 100 bp
* 135563 139847: contig of 4285 bp in length
* 139848 139947: gap of 100 bp
* 139948 146731: contig of 6784 bp in length
* 146732 146831: gap of 100 bp
* 146832 151469: contig of 4638 bp in length
* 151470 151569: gap of 100 bp
* 151570 159901: contig of 8332 bp in length
* 159902 160001: gap of 100 bp
* 160002 168308: contig of 8307 bp in length
* 168309 168408: gap of 100 bp
* 168409 176195: contig of 7787 bp in length
* 176196 176295: gap of 100 bp
* 176296 189303: contig of 13008 bp in length
* 189304 189403: gap of 100 bp
* 189404 202135: contig of 12732 bp in length
* 202136 202235: gap of 100 bp
* 202236 221365: contig of 19130 bp in length.

FEATURES
      Location/Qualifiers
      1..221365
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="RP11-462N18"
      /clone_lib="RPC1-11 Human Male BAC"
      1..1108
      /note="assembly_fragment"
      misc_feature
      1209..2292
      /note="assembly_fragment"
      misc_feature
      2393..3673
      /note="assembly_fragment"
      misc_feature
      3774..4947
      /note="assembly_fragment"
      misc_feature
      5048..6104
      /note="assembly_fragment"
      misc_feature
      6205..7259
      /note="assembly_fragment"
      misc_feature
      7360..8667

alignment_scores:
      Quality: 629.50      Length: 476
      Ratio: 3.459      Gaps: 6
      Percent Similarity: 38.235      Percent Identity: 34.244

alignment_block:
US-09-135-238B-2_COPY_18_253 x AC023534 ..
Align seg 1/1 to: AC023534 from: 1 to: 221365
      1 Argf1eLeuProGluValysValGluGlyGluLeuGlyGlySerValth 17
      |||
33190 AGATCTCCCAAGTAAGGTAGAGGGGAGCTGGCGGATCATCTTAC 33239
      |||
17 r1eLysCysProLeuProGluMethisValArgilletyrLeuCysArg 34
      |||
33240 CATCAAGTCCCACTTCCCTGAATGCATGTGAGGATATATCTGTCCGG 33289
      |||
34 luMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsnPhe 50
      |||
33290 AGATGGCTGGATCTGGAACATGTGTACGTGTATCCACCAACTTC 33339
      |||
51 l1eLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProArgly 67
```

```
|||||
33340 ATCAGGCAGAAATACAGGCGCGATTACTCTGAGCAATATCCACGGAA 33389
67 sAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerglyV 84
|||||
33390 GAATCTGTCTCTAGTGGAGGTAAACACAGCTGACAGAAAGTGACAGCGGAG 33439
84 aLtyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrClnLys 100
|||||
33440 TCTATGCCCTGCGGAGCGGCGATGAACACAGACCGGGGAAAGACCCAGAAA 33489
101 ValThrLeuAsnValHis..... 106
|||||
33490 GTCACCCGTAATGTGCACAGCTGTGTAGTTCCTCGCTGATTGGAGGCTCA 33539
106 ..... 106
33540 GACTACCCAGAAATGTTTCACTTGGGAAGATAGGAACAGACATCACTAACT 33589
106 ..... 106
33590 GTACAGAGTCTTAGACTCTTACACTGTGGCCATATAAGCCACATATGA 33639
106 ..... 106
33640 TTGAGATCCATATGAACCTCATAGTCAATTACCTCCATGCTGCCACAAA 33689
106 ..... 106
33690 TTTTACCCAAACATGTATGAAAAGAACAGATAGGTGCTCTTTGTAT 33739
106 ..... 106
33740 TGTTCATCTCCAGCGCATGAGGGAGTAGAGGCTTGAGATTGAGGAG 33789
106 ..... 106
33790 AGGCTCTGATGAGGAATTTGGAGCAAAAGGAATAACTATCCAAAG 33839
106 ..... 106
33840 TAATTGTCCTATGAGAAATAAGGAAGACACTAAATTATCCAGCCACAT 33889
106 ..... 106
33890 TTATAGGCCAACTATTACTCTGTGACCTGATCCCTTATTTCTCGAA 33939
106 ..... 106
33940 CCCCCTTTCTCCTTCTCCACTGAGTTCCTCCACACATATGCAGCAGT 33989
106 ..... 106
33990 AATTCAACAGCTAGATCCCACTGCTGCTGTTTGCAGGGAATG 34039
106 ..... 106
34040 TTGACTGGACTAGGGAAACAGTATTTCCTCCAGGAAGCCCAACCCA 34089
106 ..... 106
34090 GCTCCTTCTGCTTCTTTCAGTGAGCCCTGCGCTCATCCCTCACTGTC 34139
106 ..... 106
34140 ATCCACTTTGCTGAATTAGCTTCAGTTTACACCAAGCCTCAGTCTCCC 34189
107 ..... Ser..GluTyrGluProS 112
34190 CCTAGAGACCACCAACATATCTCTTCTTCCAAATAGAAATACGAGCAT 34239
112 eTTPGluGluProMetProGluThrProLysTrpPheHisLeuPro 128
|||||
```

```
34240 CATGGGAAGCAGCCCAATGCTGAGACTCCAAAATGGTTTCATCTGCC 34289
129 TyrLeuPheClnMetProAlaTyrAlaSerSerLysPheValThrTr 145
|||||
34290 TATTGTTCAGATGCTGCATATGCCAGTCTTCCAAATTCGTAACCAAG 34339
145 g...ValThrThrProAlaGlnArgGlyLysValPro..... 156
|||||
34340 AGGTCAGTTCACCATAGCAGGGGTAGAGAGGTGGGTGAGATGCTGTGCC 34389
157 .....ProValHisHisSerSerPro 163
34390 AATGTGAAAATGAGGAGGGGCTGTATGTTCAAGTTCA..... 34427
164 ThrThrClnIleThrHisArgProArgValSerArgAlaSerSerValal 180
|||||
34428 .....AGAGTGGCTGGCAGAGCTGAGCTGAGTCTTAGGTGTGTATTTCCC 34471
180 aGlyAspLysProArgThrPheLeuProSerThrThrAlaSerLysIleS 197
|||||
34472 TGCCTTTAGAGAGGCGACATACCTCCCTGAGATCTGCGGTCATATCCTT 34521
197 eAlaLeuGluGlyLeuLysProGlnThrProSerTyrAsnHisHis 213
|||||
34522 CAAGCCTGTCGACAGCTCAGTCCCTGCTGCTGCTGTCAGAACTTACAT 34571
214 .....ThrArgLeuHisArgGlnArg 220
34572 ATTACCAACAGGTGGCACCACACAGCAA 34598
seq_name: gb_htg22:AC027719
seq_documentation_block:
LOCUS AC027719 58628 bp DNA HTG 01-APR-2000
DEFINITION Homo sapiens chromosome 8 clone RP11-258P9 map 8, LOW-PASS SEQUENCE
SAMPLING.
ACCESSION AC027719.1 GI:7382531
VERSION HTG: HTGS_PHASE0.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 58628)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 8, clone RP11-258P9
JOURNAL Unpublished
AUTHORS
2 (bases 1 to 58628)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,P., Fitzhugh,W., Gage,D.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessfaye,S., Theodore,J., Tirrelli,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
JOURNAL Submitted (01-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT All repeats were identified using RepeatMasker:
```

Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L7449  
Center clone name: 258\_P\_9  
-----

\* NOTE: This record contains 74 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 671: contig of 671 bp in length  
\* 672 771: gap of 100 bp  
\* 772 1469: contig of 698 bp in length  
\* 1470 1569: gap of 100 bp  
\* 1570 2261: contig of 692 bp in length  
\* 2262 2361: gap of 100 bp  
\* 2362 3051: contig of 690 bp in length  
\* 3052 3151: gap of 100 bp  
\* 3152 3816: contig of 665 bp in length  
\* 3817 3916: gap of 100 bp  
\* 3917 4597: contig of 681 bp in length  
\* 4598 4697: gap of 100 bp  
\* 4698 5395: contig of 698 bp in length  
\* 5396 5495: gap of 100 bp  
\* 5496 6185: contig of 690 bp in length  
\* 6186 6285: gap of 100 bp  
\* 6286 6969: contig of 684 bp in length  
\* 6970 7069: gap of 100 bp  
\* 7070 7780: contig of 711 bp in length  
\* 7781 7880: gap of 100 bp  
\* 7881 8590: contig of 710 bp in length  
\* 8591 8690: gap of 100 bp  
\* 8691 9372: contig of 682 bp in length  
\* 9373 9472: gap of 100 bp  
\* 9473 10162: contig of 690 bp in length  
\* 10163 10262: gap of 100 bp  
\* 10263 10944: contig of 682 bp in length  
\* 10945 11044: gap of 100 bp  
\* 11045 11747: contig of 703 bp in length  
\* 11748 11847: gap of 100 bp  
\* 11848 12547: contig of 700 bp in length  
\* 12548 12647: gap of 100 bp  
\* 12648 13336: contig of 689 bp in length  
\* 13337 13436: gap of 100 bp  
\* 13437 14103: contig of 667 bp in length  
\* 14104 14203: gap of 100 bp  
\* 14204 14872: contig of 669 bp in length  
\* 14873 14972: gap of 100 bp  
\* 14973 15682: contig of 710 bp in length  
\* 15683 15782: gap of 100 bp  
\* 15783 16491: contig of 709 bp in length  
\* 16492 16591: gap of 100 bp  
\* 16592 17290: contig of 699 bp in length  
\* 17291 17390: gap of 100 bp  
\* 17391 18088: contig of 698 bp in length  
\* 18089 18188: gap of 100 bp  
\* 18189 18883: contig of 695 bp in length  
\* 18884 18983: gap of 100 bp  
\* 18984 19671: contig of 688 bp in length  
\* 19672 19771: gap of 100 bp  
\* 19772 20413: contig of 642 bp in length

\* 20414 20513: gap of 100 bp  
\* 20514 21196: contig of 683 bp in length  
\* 21197 21296: gap of 100 bp  
\* 21297 21992: contig of 696 bp in length  
\* 21993 22092: gap of 100 bp  
\* 22093 22804: contig of 712 bp in length  
\* 22805 22904: gap of 100 bp  
\* 22905 23604: contig of 700 bp in length  
\* 23605 23704: gap of 100 bp  
\* 23705 24402: contig of 698 bp in length  
\* 24403 24502: gap of 100 bp  
\* 24503 25191: contig of 689 bp in length  
\* 25192 25291: gap of 100 bp  
\* 25292 25974: contig of 683 bp in length  
\* 25975 26074: gap of 100 bp  
\* 26075 26769: contig of 695 bp in length  
\* 26770 26869: gap of 100 bp  
\* 26870 27572: contig of 703 bp in length  
\* 27573 27672: gap of 100 bp  
\* 27673 28373: contig of 701 bp in length  
\* 28374 28473: gap of 100 bp  
\* 28474 29165: contig of 692 bp in length  
\* 29166 29265: gap of 100 bp  
\* 29266 29956: contig of 691 bp in length  
\* 29957 30056: gap of 100 bp  
\* 30057 30741: contig of 685 bp in length  
\* 30742 30841: gap of 100 bp  
\* 30842 31519: contig of 678 bp in length  
\* 31520 31619: gap of 100 bp  
\* 31620 32323: contig of 704 bp in length  
\* 32324 32423: gap of 100 bp  
\* 32424 33110: contig of 687 bp in length  
\* 33111 33210: gap of 100 bp  
\* 33211 33902: contig of 692 bp in length  
\* 33903 34002: gap of 100 bp  
\* 34003 34708: contig of 706 bp in length  
\* 34709 34808: gap of 100 bp  
\* 34809 35500: contig of 692 bp in length  
\* 35501 35600: gap of 100 bp  
\* 35601 36300: contig of 700 bp in length  
\* 36301 36400: gap of 100 bp  
\* 36401 37097: contig of 697 bp in length  
\* 37098 37197: gap of 100 bp  
\* 37198 37900: contig of 703 bp in length  
\* 37901 38000: gap of 100 bp  
\* 38001 38695: contig of 695 bp in length  
\* 38696 38795: gap of 100 bp  
\* 38796 39487: contig of 692 bp in length  
\* 39488 39587: gap of 100 bp  
\* 39588 40291: contig of 704 bp in length  
\* 40292 40391: gap of 100 bp  
\* 40392 41101: contig of 710 bp in length  
\* 41102 41201: gap of 100 bp  
\* 41202 41886: contig of 685 bp in length  
\* 41887 41986: gap of 100 bp  
\* 41987 42686: contig of 700 bp in length  
\* 42687 42786: gap of 100 bp  
\* 42787 43484: contig of 698 bp in length  
\* 43485 43584: gap of 100 bp  
\* 43585 44276: contig of 692 bp in length  
\* 44277 44376: gap of 100 bp  
\* 44377 45078: contig of 702 bp in length  
\* 45079 45178: gap of 100 bp  
\* 45179 45869: contig of 691 bp in length  
\* 45870 45969: gap of 100 bp  
\* 45970 46664: contig of 695 bp in length  
\* 46665 46764: gap of 100 bp  
\* 46765 47456: contig of 692 bp in length  
\* 47457 47556: gap of 100 bp  
\* 47557 48248: contig of 693 bp in length  
\* 48250 48349: gap of 100 bp  
\* 48350 49042: contig of 693 bp in length  
\* 49043 49142: gap of 100 bp

```

* 49143 49843: contig of 701 bp in length
* 49844 49943: gap of 100 bp
* 49944 50640: contig of 697 bp in length
* 50641 50740: gap of 100 bp
* 50741 51438: contig of 698 bp in length
* 51439 51538: gap of 100 bp
* 51539 52331: contig of 693 bp in length
* 52332 52331: gap of 100 bp
* 52332 53020: contig of 689 bp in length
* 53021 53120: gap of 100 bp
* 53121 53857: contig of 737 bp in length
* 53858 53957: gap of 100 bp
* 53958 54652: contig of 695 bp in length
* 54653 54752: gap of 100 bp

```

## alignment\_scores:

```

Quality: 252.00 Length: 48
Ratio: 5.250 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

## alignment\_block:

```

US-09-135-238B-2_COPY_18_253 x AC027719/rev ..

```

```

Align seg 1/1 to reverse of: AC027719 from: 1 to: 58628

```

```

1 ArgileLeuProGluValLysValGluGlyLeuGlySerValTh 17
|||||
40580 AGGATCTCCAGAACTAAGGTAGAGGGGAGCTGGGGATCAGTTAC 40531
|||||
17 rileyCysProLeuProGluMetHisValArgIleTyLeuCysArg 34
|||||
40530 CATCAAGTGGCCACTCTCTGAAATGCATGTGAGATATATCTGTCGGG 40481
|||||
34 luMetAlaGlySerGlyThrCysGlyThrValValSerThrThr 48
|||||
40480 AGATGGCTGGATCTGGACATGTGGTACCGTGTATCCACCACC 40437
|||||

```

```

seq_name: gb_ro:MMU06431

```

```

seq_documentation_block:
LOCUS MMU06431 3095 bp mRNA 26-MAY-1995
DEFINITION Mus musculus B6/CBAF1J polymeric immunoglobulin receptor mRNA,
complete cds.
ACCESSION U06431
VERSION U06431.1 GI:458421
KEYWORDS .
SOURCE mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 3095)
AUTHORS Piskurich,J.F., Blanchard,M.H., Youngman,K.R., France,J.A. and
Kaetzel,C.S.
TITLE Molecular cloning of the mouse polymeric Ig receptor. Functional
regions of the molecule are conserved among five mammalian species
J. Immunol. 154 (4), 1735-1747 (1995)
JOURNAL 95138517
MEDLINE 2 (sites)
REFERENCE Piskurich,J.F., Hsieh,M., Youngman,K.R. and Kaetzel,C.S.
AUTHORS Molecular cloning of mouse polymeric Ig receptor
TITLE J. Immunol. 150, 38A-38A (1993)
JOURNAL
REFERENCE 3 (bases 1 to 3095)
AUTHORS Piskurich,J.F.
TITLE Molecular Cloning and Regulation of the Polymeric Immunoglobulin
Receptor
JOURNAL Thesis (1994) Pathology, Case Western Reserve University
REFERENCE 4 (bases 1 to 3095)
AUTHORS Piskurich,J.F.
TITLE Direct Submission
JOURNAL Submitted (08-FEB-1994) Janet F. Piskurich, Case Western Reserve
University, Pathology, Biomedical Research Building, Cleveland, OH
44060, USA
FEATURES Location/Qualifiers

```

```

source
1. .3095
/organism="Mus musculus"
/strain="B6/CBAF1J (C57 Black 6 x CBA)"
/db_xref="taxon:10090"
/clone="G20"
/clone_lib="catalog #935302, Stratagene Cloning Systems,
La Jolla, CA"
/sex="female"
/tissue_type="liver"
<1. .84
85. .138
85. .2400
/codon_start=1
/function="binds and transports polymeric immunoglobulin"
/product="polymeric immunoglobulin receptor"
/protein_id="AA67440.1"
/db_xref="GI:458422"
/translation="MRLYLFTLLTVFSGVSTKSPIFGPQEVSSIEGDSVSICTYYPD
TSVNRHTRKYWCROGAGMCTTLISSGYSKEYSGRANLINFPENNTFVINTQLTQ
DPTGSKGCLGTSNRGLSDVSEVNPQYVPELPSDTHVTYKIDGRNVTIECPFKRENVP
SKKSLCKTNGSCLEVIDSTKVNPSYIGRAKLFMKGTDLTVYVNIHLTHNDAGLY
ICQAGEPSADKNVDLOVLAPEPELLYKDLRSVTFECDLGREVANEAKVLCRMNKE
TCDVIINTLGRKDPDFEGRILITPKDDNGRFSVLTGLRKEADAGVOCGASHSSLPQE
GWPLOTWOLFVNESTIPNRRSVKVGTSVALACPNKSSSLAKYKWRGDCNG
HCPALYGTQAOQVEYEGRLALFDOPGNGYTVILLQLTDEAGFYWCLNGDSRWHT
TYELQVAEATREPENLVITFONATAVIGETFTVTSCHTPCAFYISQEKYCKWSNGSHIL
PSHDEGARQSSVCDSSQLVSKMLNPVSRKEDEGWYCWGKGQGTGYETATYIAVEE
RTRGSHVNPDTANAKRAVLEEVDSSISEKENKAIPNPGFFANEREIIONYRDAQ
ENRAGSDAGSADGQSRSSSKVLFSTLPLGLVLAYGAIAVWVAVRHRKKNVDRMSIS
SYRTDISMADFKNSDLGGNDNMGASPDTOQTIEGKDEIVTTTCTAEPEESKAKR
SSKEADMAVSFAFLLOSSTIAAQVHDGPOEA"
mat_peptide 139. .2397
misc_feature 139. .2019
misc_feature 139. .2088
misc_feature 2089. .2397
misc_feature 2401. .3095
3'UTR 861 a 796 c 784 g 654 t
BASE COUNT
ORIGIN
alignment_scores:
Quality: 197.00 Length: 148
Ratio: 2.165 Gaps: 5
Percent Similarity: 61.486 Percent Identity: 32.432
alignment_block:
US-09-135-238B-2_COPY_18_253 x MMU06431 ..
Align seg 1/1 to: MMU06431 from: 1 to: 3095
13 GlyGlySerValThrIleLysCysProLeuProGlu.....Me 25
||| |||||:|||| ||| |||:
181 GCGGACTCTGTTCCATCAGCTGCTACTACCAGACACCTCTGTCAACCG 230
25 tHisValArgIleTyLeuCysArgGluMetAlaGlySerGlyThrCysG 42
||||| ||| |||||: ||| ||||| |||
231 GCACACCCGGAATACTGGTCCGACAGAGAGCC...AGCGGCATGTCCA 277
42 lyThrValValSerThrThrAsnPhelIleLysAlaGluTyLysGlyArg 58
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
278 CAACGCTCATCTTCAATATGGCTACCTCTCCAAAGGAGTATTTCAGGCAGA 327
59 ValThrLeuLysGlnTyProArgLysAsnLeuPheLeuValGluValTh 75
:|||||: |||||: |||||: |||||: |||||: |||||: |||||:
328 GCCAACCTCATCACTCCCGACAGAACACACATTTGTGATTAAACATTGA 377
75 rGlnLeuThrGluSerAspSerClyValTyAlaCysGlyAlaGlyMeta 92
|||||:|||||:|||||: ||| ||||| |||
378 GCAGCTCACCCAGGACGACACTGGGAGCTACAGTGTGGCTGGGT...A 424

```





REFERENCE 2 (bases 1 to 417)  
 AUTHORS Martin,M.G., Gutierrez,E.M., Lam,J.T., Li,T.W.H. and Wang,J.  
 TITLE Direct Submission  
 JOURNAL Submitted (31-DEC-1996) Pediatrics, UCLA School of Medicine, 10833 Le Conte Ave, Los Angeles, CA 90095, USA

FEATURES  
 source Location/Qualifiers  
 1. .417  
 /organism="Mus musculus"  
 /strain="129SVJ"  
 /db\_xref="taxon:10090"  
 43. .387  
 /gene="pIgr"  
 /number=3

exon 105 a 120 c 102 g 90 t

BASE COUNT 105 a 120 c 102 g 90 t  
 ORIGIN

alignment\_scores:  
 Quality: 188.00 Length: 97  
 Ratio: 2.648 Gaps: 3  
 Percent Similarity: 73.196 Percent Identity: 40.206

alignment\_block:  
 US-09-135-238b-2\_copy\_18\_253 x MMPIGR03 ..

Align seg 1/1 to: MMPIGR03 from: 1 to: 417

13 GlyGlySerValThrIleLysCysProLeuProGlu.....Me 25  
 ||| |||||:|||| ||| |||:  
 96 GGGGACTCTGTTTCCATCAGTGCTACTACCCAGACACCTCTGTCAACCG 145  
 25 tHISValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 42  
 |||||:|||| ||| |||||: ||| ||||| |||  
 146 GCACACCCGGAAATCTGGTGGCCGACAAAGAGCC...AGCGCATGTGCA 192  
 42 lyThrValValSerThrThrAsnPhelLysAlaGluTyrLysGlyArg 58  
 |||||:|||| ||| |||||: ||| ||||| |||  
 193 CAACGGTCATCTTCAATGGTCTACTCTCAAGGAGTATTCAGGCAGA 242  
 59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75  
 :|||:|||| ||| |||||: ||| |||||: ||| |||||: ||| |||||:  
 243 GCCACCTCATCACTTCCAGAGACACACACATTTGTGATTACATGTA 292  
 75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 92  
 |||||:|||| ||| |||||: ||| ||||| |||  
 293 GCAGCTCACCCAGACGACACTGGGAGCTACAAGTGTGGCTGGGT...A 339  
 92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnVal 105  
 :|||:|||| ||| |||||: ||| ||||| |||  
 340 CCAGTAACCGAGGCGCTGCTCTCGATGTCAGCCTGGAGGTC 380

seq\_name: gb\_ro:MMPIGR3

seq\_documentation\_block:  
 LOCUS MMPIGR3 505 bp DNA ROD 09-APR-1998  
 DEFINITION Mus musculus pIgr gene, exon 3.  
 ACCESSION Y16525  
 VERSION Y16525.1 GI:3046750  
 KEYWORDS pIgr gene; polymeric immunoglobulin receptor.  
 SOURCE house mouse.

ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
 Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 505)  
 AUTHORS de Groot,N., Vollebregt,E., Lee,S.H., Verbeet,M.P. and de Boer,H.A.  
 TITLE Molecular cloning and exon-intron organization of the gene encoding the murine polymeric immunoglobulin receptor

JOURNAL Unpublished  
 2 (bases 1 to 505)  
 AUTHORS de Groot,N.  
 TITLE Direct Submission  
 JOURNAL Submitted (11-FEB-1998) N. De Groot, Medical Biotechnology Dept., Leiden University, P.O. box 9502, 2300 RA Leiden, NETHERLANDS

FEATURES  
 source Location/Qualifiers  
 1. .505  
 /organism="Mus musculus"  
 /strain="129"  
 /db\_xref="taxon:10090"  
 /tissue\_type="liver"  
 77. .421  
 /gene="pIgr"  
 77. .421  
 /gene="pIgr"  
 /label-ex3  
 /usedin=Y16524:pig\_cds  
 /usedin=Y16524:pig\_mrna

BASE COUNT 130 a 136 c 127 g 112 t  
 ORIGIN

alignment\_scores:  
 Quality: 188.00 Length: 97  
 Ratio: 2.648 Gaps: 3  
 Percent Similarity: 73.196 Percent Identity: 40.206

alignment\_block:  
 US-09-135-238b-2\_copy\_18\_253 x MMPIGR3 ..

Align seg 1/1 to: MMPIGR3 from: 1 to: 505

13 GlyGlySerValThrIleLysCysProLeuProGlu.....Me 25  
 ||| |||||:|||| ||| |||:  
 130 GGGGACTCTGTTTCCATCAGTGCTACTACCCAGACACCTCTGTCAACCG 179  
 25 tHISValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 42  
 |||||:|||| ||| |||||: ||| ||||| |||  
 180 GCACACCCGGAAATCTGGTGGCCGACAAAGAGCC...AGCGCATGTGCA 226  
 42 lyThrValValSerThrThrAsnPhelLysAlaGluTyrLysGlyArg 58  
 |||||:|||| ||| |||||: ||| ||||| |||  
 227 CAACGGTCATCTTCAATGGTCTACTCTCAAGGAGTATTCAGGCAGA 276  
 59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75  
 :|||:|||| ||| |||||: ||| |||||: ||| |||||: ||| |||||:  
 277 GCCACCTCATCACTTCCAGAGACACACATTTGTGATTACATGTA 326  
 75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 92  
 |||||:|||| ||| |||||: ||| ||||| |||  
 327 GCAGCTCACCCAGACGACACTGGGAGCTACAAGTGTGGCTGGGT...A 373  
 92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnVal 105  
 :|||:|||| ||| |||||: ||| ||||| |||  
 374 CCAGTAACCGAGGCGCTGCTCTCGATGTCAGCCTGGAGGTC 414

seq\_name: gb\_pat:E15253

seq\_documentation\_block:  
 LOCUS E15253 29392 bp DNA PAT 28-JUL-1999  
 DEFINITION Mouse pIgr receptor gene.  
 ACCESSION E15253  
 VERSION E15253.1 GI:5709936  
 KEYWORDS JP 1998057066-A/1.  
 SOURCE Mus sp.  
 ORGANISM Mus sp.

Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 29392)  
 AUTHORS Sato,T. and Kushi,A.  
 TITLE MURINE PIG RECEPTOR PROTEIN GENE  
 JOURNAL Patent: JP 1998057066-A 03-MAR-1998;  
 YAKULT HONSHA CO LTD

COMMENT OS Mus sp. (mouse)  
 PN JP 1998057066-A/1  
 PD 03-MAR-1998  
 PF 19-AUG-1996 JP 1996217154  
 PI SATO TAKASHI, KUSHIRO AKIRA

PC C12N15/09,C07H21/04,C07K14/705;  
CC strandedness: Double;  
CC topology: Linear;  
FH Key Location/Qualifiers  
FT source 1..29392  
FT /organism="Mus sp."  
FT /clone\_lib="genomic DNA from female 129SVJ FT  
FT mouse liver in  
FT lambda fix vector'.  
FEATURES  
source Location/Qualifiers  
1..29392  
/organism="Mus sp."  
/db\_xref="taxon:10095"  
BASE COUNT 8318 a 6747 c 6514 g 7813 t  
ORIGIN  
  
alignment\_scores:  
Quality: 188.00 Length: 97  
Ratio: 2.648 Gaps: 3  
Percent Similarity: 73.196 Percent Identity: 40.206  
  
alignment\_block:  
US-09-135-238B-2\_COPY\_18\_253 x EI5253 ..  
  
Align seg 1/1 to: EI5253 from: 1 to: 29392  
  
13 GlyGlySerValThrIleLysCysProLeuProGlu.....Me 25  
||| |||||:|||| ||| |||||  
18245 GGGACTCTGTTCCATCAGTGCTACTACCCAGACACCTCTGTCAACCG 18294  
  
25 thisValArgIleTyLeuCysArgGluMetAlaGlySerGlyThrCysG 42  
|||||:|||| ||| |||||:|||| ||| |||||:||||  
18295 GCACACCCGGAATACTGTGCGCAGCAGGAGCC...ACGGCAGTGTGA 18341  
  
42 lyThrValValSerThrThrAsnPhleIleLysAlaGluTyLysGlyArg 58  
|||||:||||:||||:||||:||||: |||||:|||||  
18342 CAACGCTCATCTCTTCAAATGCTACTCTCCAGGAGTATTTCAGGCAGA 18391  
  
59 valThrLeuLysGlnTyProArgLysAsnLeuPheLeuValGluValth 75  
:||||:|||| :||||:||||:|||| :||||:||||: |||||  
18392 GCAACCTCATCACTCCAGACACACACATTGTGATTAAACATTGA 18441  
  
75 rGlnLeuThrGluSerAspSerGlyValTyraLacCysGlyAlaGlyMeta 92  
|||||:||||:||||:||||: ||| |||||:||||  
18442 GCAGCTACCCAGGAGCAGCAGCTGAGGAGCTACAGTGTGGCCTGGGT...A 18488  
  
92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnVal 105  
:||||:||||| :||| |||||:|||||  
18489 CCAGTAACCCAGGAGGCTGCTCTTCGATGTCAGCCTGGAGGTC 18529  
  
seq\_name: gb\_ro:AB001489  
  
seq\_documentation\_block:  
LOCUS AB001489 29392 bp DNA ROD 14-APR-2000  
DEFINITION Mus musculus DNA for polyimmunoglobulin receptor, complete cds.  
ACCESSION AB001489  
VERSION AB001489.1 GI:2804245  
KEYWORDS polyimmunoglobulin receptor.  
SOURCE Mus musculus (strain:129SVJ) female liver DNA.  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
REFERENCE  
1 (sites)  
AUTHORS kushiro,A. and Sato,T.  
TITLE PolymERIC immunoglobulin receptor gene of mouse: sequence,  
structure and chromosomal location  
JOURNAL Gene 204 (1-2), 277-282 (1997)  
MEDLINE 98094283  
REFERENCE 2 (bases 1 to 29392)  
AUTHORS Sato,T.  
TITLE Direct Submission

JOURNAL Submitted (26-FEB-1997) to the DDBJ/EMBL/GenBank databases. Takashi  
Sato, Yakult Central Institute for Microbiological Research,  
Culture Collection Center; 1796 Yaho, Kunitachi-shi, Tokyo 186,  
Japan (Tel:425-77-8960, Fax:425-77-3020)  
FEATURES  
Location/Qualifiers  
source 1..29392  
/organism="Mus musculus"  
/strain="129SVJ"  
/db\_xref="taxon:10090"  
/chromosome="1"  
/map="1E3"  
/sex="female"  
/tissue\_type="liver"  
CAAT\_signal 2470..2474  
TATA\_signal 2580..2585  
CAAT\_signal 2878..2882  
CAAT\_signal 3176..3180  
CAAT\_signal 3251..3255  
TATA\_signal 3350..3355  
5'UTR 3389..3501  
/note="based on the putative transcription start"  
exon 3389..3501  
/number=1  
intron 3502..11067  
/number=1  
gene 11124..27591  
/gene="pigr"  
join(11124..11166,18192..18536,21130..21783,23022..23354,  
23908..24240,25133..25325,25686..25813,26183..26314,  
26426..26484,27496..27591)  
/gene="pigr"  
/function="mediates secretion of IgA and IgM"  
/codon\_start=1  
/product="polyimmunoglobulin receptor"  
/protein\_id="BAA24431.1"  
/db\_xref="GI:2804246"  
/translation="MRLYLFTLLVTVFSGVTKSPFGPQWVSSTEGDSVSICTCYPD  
TSVNRHTRKYMCROGAGCMCTLISSNGYLSKEYSGRANLINFENNTFVINIQLTQ  
DDTSYKCGLCGTSNGLSFDVLSVQVPELPDTHYTKDIGRNVTECFKRENAP  
SKKSKKTKNSCELVIDSTERKNPSYIGRAKLFMKGTDLTFVYNISHLTHNDAGLY  
ICQAGSPSADKKNDVQLAPPEPELKYDLRSVTFECDLGRVANEAKYLCRMNKE  
TCQVINTLGKRPDPFEGRLITPKDDNGRFSVLITGLRKEDAGHYOCGAHSSGLPQE  
GWPIOTWOLFVNEESTIPNRSVVKVTGGSVAIACPNPKESSLKYWCWEGSDGNG  
HCPVLVGTQAOVQEEYEGRLALFDQPGNGTYTVILNOLTTEDAGFYWCLTNGDSRWRT  
TIELQVATREPNLEVTQPNATVILGETFTFVSCHYPCFKFYQSKYCKWKNKSGCHIL  
FSDHGAQRQSVSCDQSSQLVSMTLNPNVSKDEGWYWCVKQOGTYGETTAIYIAVEE  
RTRGSHVNPDTANARAKVALEEEVYDSSISEKENKAIPNPGFANEREIQNVGDQAO  
ENRAGSDAGSADGSRSSKVLFTSLVPLGLVLAAGIAVWVARVHRKKNVDRMSIS  
SVRTDISMADFKNRDLGGNDNMGASPTQQTQVTIEGKDEIVTTTCTAEPESKAKR  
SSKEADNAYSAFLQSSSTIAAQVHDGPQEA"  
11124..11166  
/gene="pigr"  
/number=2  
intron 11167..18191  
/gene="pigr"  
/number=2  
exon 18192..18536  
/gene="pigr"  
/number=3  
intron 18537..21129  
/gene="pigr"  
/number=3  
exon 21130..21783  
/gene="pigr"  
/number=4  
intron 21784..23021  
/gene="pigr"  
/number=4  
exon 23022..23354  
/gene="pigr"  
/number=5  
intron 23355..23907  
/gene="pigr"



```
170 GSTAACTCGGTCTCCATCATCGTGTACTACCCAGACACCTCTGTCAACGG 219
25 thisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 42
||||:||||| ||| |||||:| ||| ||||| |||:
220 GCACACCCGGAATACTGTGCGCACAAGAGCC...AACGGCTACTGG 266
42 lyThrValValSerThrThrAsnPhelIleLysAlaGluTyrLysGlyArg 58
||||:||||| |||||:||||| |||||:||||| |||||:|||||
267 CAACCTCATCTCTCAATGCTACTCTCGAAGGAGTATTCAAGCAGA 316
59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75
||||:||||| |||||:||||| |||||:||||| |||||:|||||
317 GCAGGCTCATCACTTCCAGAGATAGCACATTTGCTGATTAACATTGC 366
75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 92
||||:||||| |||||:||||| |||||:||||| |||||:|||||
367 ACATCTCACCCAGGAGACACTGGAGCTACAAGTGTGTTCTGGGT...A 413
92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 108
||||:||||| |||||:||||| |||||:||||| |||||:|||||
414 CCACAAACCGAGGCGCTGTTTTCGATGTCCAGCTGGAGTCT... 454
109 TyrGluProSerIrpGluGluGlnProMetProGluThrProLysIrpPh 125
||||:||||| |||||:||||| |||||:||||| |||||:|||||
455 ..... ACCAGGTTCTCTGAGTTCCTCAATGACAC 483
125 eHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSerSerLysP 142
||||:||||| |||||:||||| |||||:||||| |||||:|||||
484 CCAATGTC.....TACACAAAGG 500
142 heValThrArgValThrProAlaGlnArgGlyLysValProVal 158
||||:||||| |||||:||||| |||||:||||| |||||:|||||
501 ACATAGGACAGAACTGTACCAATGCGCTTCAAGAGGGAATGCT 550
159 HisHisSer.SerProThrThrGlnIleThrHisArgPro 171
||||:||||| |||||:||||| |||||:||||| |||||:|||||
551 CATAGCAGAATCCCTGTGTAGAAAGAGGAGGAGGCT 590
seq_name: gb_om:AB032195
seq_documentation_block: 2461 bp mRNA MAM 03-FEB-2000
LOCUS AB032195
DEFINITION Sus scrofa mRNA for poly-Ig receptor, complete cds.
ACCESSION AB032195
VERSION AB032195.2 GI:6863079
KEYWORDS poly-Ig receptor.
SOURCE Sus scrofa female cDNA to mRNA, clone:SC1.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 2461)
Sone,T. and Kumura,H.
TITLE Porcine mammary gland cDNA clone, similar to poly-Ig receptor
JOURNAL Published Only in DataBase (1999) In press
REFERENCE 2 (bases 1 to 2461)
Sone,T. and Kumura,H.
AUTHORS Direct Submission
TITLE Submitted (07-SEP-1999) to the DDBJ/EMBL/GenBank databases. Teruo
JOURNAL Sone, Hokkaido University, Laboratory of Applied Microbiology,
Faculty of Agriculture; Kita-9, Nishi-9, Kita-ku, Sapporo, Hokkaido
060-8589, Japan (E-mail:sonet@chem.agr.hokudai.ac.jp,
Tel:81-11-706-2502, Fax:81-11-706-4961)
COMMENT On Feb 3, 2000 this sequence version replaced gi:5881588.
Sequence updated (01-Feb-2000).
FEATURES
source
1. .2461
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone="SC1"
/sex="female"
sig_peptide
58. .111
/product="putative"
CDS
58. .2334
```

```
/codon_start=1
/product="poly-Ig receptor"
/protein_id="BAA84283.2"
/db_xref="GI:6863080"
/translation="MTRFFYACLLAIFPVSMKSPIFGPDVSSVEGSSVIRCIYPA
TSVNRHSRYWCRIGAKGRCTLIISSEGYISKDYGRANLTNFPENGTFVMDIGHLTR
GDSGLYKGLGSISSRGLSFDVLSYSGPGQGVHVYTDLGSITVTINCPFKSENAG
KPKSVYKLGQIRLVIDSNGVNLNNFTNRAHLSIQGTNOLVFSVINRIOLRGAGIY
ICQADESSADLOVLKPEPELIGDLRGSVTFDCALGOEMANVAKFLCOLKNGKTCN
VVIITLGRKAQDFEGRIILITPKNSHFVSHITGLKEDAGHYLAGAHPDGPKEGMPV
QAWQLFINEDTIMPPRSSVVRGVGSVAVTCPIPNKETSLSKIKYWCWREBENGRCPO
LIVESGLVNEQYEGRLALYEENGTFTVILNLTNRDAGFYCLTNEDSDSRWSTVEL
KIVEGEPNLKLPENVTAWGTGTFKLSCHFPKIFYSYQIKYCKWNSITGRALPSQDEGQ
SOAFVNCDDKKSQIISLNLNPKRDEGWCVKDLGHYGETGAVYVAVEQKAGSGD
ARLANAAPAEADAIEPRARENEVLDPSSLIAEDRAVKDGEPSAGSGVPAAPGSGD
VGQGGSKRAVSTLPLVLMYLVAVTIGVLRHKKNVDRISIRSYTDLSDMSDFES
SRDFGHDNAGATLDTOETSLGKDKRFTTTTNTMEDKEPRRKRKSSKEADKATFTT
LLQANSIIAATQNGPRE"
BASE COUNT 608 a 699 c 708 g 446 t
ORIGIN
alignment_scores:
Quality: 181.00 Length: 118
Ratio: 2.207 Gaps: 4
Percent Similarity: 69.492 Percent Identity: 33.898
alignment_block:
US-09-135-238B-2_COPY_18_253 x AB032195 ..
Align seg 1/1 to: AB032195 from: 1 to: 2461
2 IleLeuProGluValLysValGluGlyGluLeu..... 12
||||:||||| |||||:||||| |||||:||||| |||||:|||||
91 ATCTCCCGAGTGGTCTCCATGAGAGTCCCATATTCGTCGCCAGGATGT 140
13 .....GlyGlySerValThrIleLysCysProLeuPro.... 23
||||:||||| |||||:||||| |||||:||||| |||||:|||||
141 GAGCAGCGTGGAAAGCAGCTCGTGTCTCCATCAGATGCTACTACCAGCCA 190
24 .....GluMetHisValArgIleTyrLeuCysArgGluMetAlaGly 37
::: ||| ||| ||| ||||| |||||:|||||
191 CCTCGCTCAACCGGCAATTCCTCGGAAGTACTGCTGCCGA...ATAGGAGCC 237
38 SerGlyThrCysGlyThrValValSerThrThrAsnPhelIleLysAlaG 54
||||| ||| |||||:||||| |||||:||||| |||||:|||||
238 AAGGCGCGCTGCACACCTCATCTCTCGGAGGCGCTACATCTCCAAGGA 287
54 uTyrLysGlyArgValThrLeuLysGlnTyrProArgLysAsnLeuPheL 71
||||| |||||:||||| |||||:||||| |||||:||||| |||||:|||||
288 CTACAAGGCGCAGAGCCACCTCACCAACTTCCAGAGACGCGACCTTCG 337
71 euValGluValThrGlnLeuThrGluSerAspSerGlyValTyrAlaCys 87
||||:||||| |||||:||||| |||||:||||| |||||:||||| |||||:|||||
338 TGATGGACATTGGCCACCTGACCCCGGCTGACTCTGGGCTCTACAAGTGT 387
88 GValAlaGlyMetAsnThrAspArgGlyLysThrGlnLysValThrLeuAs 104
||| |||||:||||| |||||:||||| |||||:||||| |||||:|||||
388 GGTCTGGCATATGACAGC...CGAGGCTGTCTTTTGTGATGTGAGCCTGGA 434
104 nVal 105
::|||
435 GGTC 438
seq_name: gb_pr2:S43435S03
seq_documentation_block: 345 bp DNA PRI 08-MAY-1993
LOCUS S43435S03
DEFINITION transmembrane secretory component [human, leukocytes, Genomic, 345
nt, segment 3 of 11].
ACCESSION S43441
VERSION S43441.1 GI:255089
KEYWORDS
```



mat\_peptide 111..2015  
BASE COUNT 477 a 540 c 607 g 407 t  
ORIGIN

alignment\_scores:  
Quality: 180.50 Length: 106  
Ratio: 2.314 Gaps: 4  
Percent Similarity: 73.585 Percent Identity: 38.679

alignment\_block:  
US-09-135-238B-2\_COPY\_18\_253 x A52093 ..

Align seg 1/1 to: A52093 from: 1 to: 2031

4 ProGluValLysValGluGlyLeuGlyGlySerValThrIleLysCy 20  
||||| :|||:||||| :|||:|||||:||||| :||  
183 CCCGAG...GAGTGAATAGTGTGAAGTAACACAGTGTCCATCACGTG 229  
20 sProLeuPro.....GluMetHisValArgIleTyrLeuCysA 33  
| ||| :||| :|||:||||| :|||:||||| :|||  
230 CTACTACCCACCCACCTCTGTCAACCGGCACACCGGAAGTACTGGTGCC 279  
33 rgGluMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsn 49  
||||| :|||:||||| :|||:|||||:||||| :|||  
280 GCGAGGAGCTAGAGTGGC...TGCATAACCTCATCTCTCGGAGGCG 326  
327 TAGCTCTCCAGCAATATGACGAGCGGTAACTCAACCTCACCACTTCCCGGA 376  
50 PheIleLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProAr 66  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
327 TAGCTCTCCAGCAATATGACGAGCGGTAACTCAACCTCACCACTTCCCGGA 376  
66 gLysAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerG 83  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
377 GAACGGCACATTGCGTGGTGAACATTGCCAGCTGAGCCAGGATGACTCCG 426  
83 lyValTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGln 99  
|| ||| ||||| |||||:|||||:|||||:||||| :|||  
427 GCGCTACAAAGTGTGGCTGGGCATCAATAGC...CGAGGCTGTGCTTT 473  
100 LysValThrLeuAsnVal 105  
|||:|||||:|||||  
474 GATGTCAGCTGGAGTGC 491

seq\_name: gb\_pat:A52091

seq\_documentation\_block:  
LOCUS A52091 2405 bp DNA PAT 11-MAR-1997  
DEFINITION Sequence 1 from Patent WO9618734.  
ACCESSION A52091  
VERSION A52091.1 GI:2304703

KEYWORDS  
SOURCE unidentified.

ORGANISM  
unclassified.

REFERENCE  
1 (bases 1 to 2405)  
AUTHORS Van,O.J., Asselbergs,F.A., Leist,C. and Gruetter,M.  
TITLE PRODUCTION OF RECOMBINANT SECRETORY COMPONENT  
JOURNAL Patent: WO 9618734-A 1 20-JUN-1996;  
CIBA GEIGY AG (CH)  
COMMENT Other publication AU 4302796 960703.

FEATURES  
Location/Qualifiers

1..2405  
/organism="unidentified"

/db\_xref="taxon:32644"

111..2405

/codon\_start=1

/product="ORIGINAL PIGR"

/protein\_id="CAA03384.1"

/db\_xref="GI:2304704"

/db\_xref="SWISS-PROT:P01833"

/translation="MLLVLTCLLAVFPFAISTKSPFPGPEVNSVEGNSVITCYPP  
TSVNRHTKRYWCROGARGGCITLISSEGYVSSKYAGRANLTNFPENCTFVVNIAQLSQ

DDSGRYKCGLGINSRGLSPDVSLEVSQGGPLLNDTKYTVTDLGRVTYINCPFKTENAQ  
KRSLYKQIGLYPVLYIDSSGYNPNYTHRIQDIOGTGQLLESVINOLRLSDAGY  
LCQAGDDSNKKKNADLOVLPKPELVEDLRGVSFTHFCALGEVANAKFLCRQSSG  
ENCDDVYNTLGGKAPAFEDGRILLNPQDKDSFVITGURKEDAGRYLCGAHSDGQLQ  
EGSPIQAWQLFVNEESTIPRSPTVKVGAGSVAVLCIPYNRKESKIKYWCLEWGAON  
GRCPLLVDSEGWKAOYEGRLSLLEFPNGTFTVILNQLTSRDAGFYWCILNIGDTLWR  
TTVEIKIIEGEPNLKVPNGVTAVLGETLKVPCHPCKFSYKRYKWCNNNTGCOALPS  
ODEGPKAFVNCDENSLVSILTNLNLTTRADEGWYGCVKOGHFYGETAAVYVAERK  
AAGSRUVSLAKADRAPEKVIDSGFREIENKATQDPRLFAEERKAVADTRQADGSRAS  
VDGSEEGQSSRALVSTLVLPIGLVLAAGAVAGVARHRNRKNVDKVSIRSTRDIS  
MSDFNSREFGANDMGASITQETSLGSGKEEFVATTESTTETKEPKKAKRSKEAE  
MAYKDFLLQSTVAAEAQDGPQA"

BASE COUNT 579 a 648 c 718 g 460 t  
ORIGIN

alignment\_scores:  
Quality: 180.50 Length: 106  
Ratio: 2.314 Gaps: 4  
Percent Similarity: 73.585 Percent Identity: 38.679

alignment\_block:  
US-09-135-238B-2\_COPY\_18\_253 x A52091 ..

Align seg 1/1 to: A52091 from: 1 to: 2405

4 ProGluValLysValGluGlyLeuGlyGlySerValThrIleLysCy 20  
||||| :|||:||||| :|||:|||||:||||| :||  
183 CCCGAG...GAGTGAATAGTGTGAAGTAACACAGTGTCCATCACGTG 229  
20 sProLeuPro.....GluMetHisValArgIleTyrLeuCysA 33  
| ||| :||| :|||:||||| :|||:||||| :|||  
230 CTACTACCCACCCACCTCTGTCAACCGGCACACCGGAAGTACTGGTGCC 279  
33 rgGluMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsn 49  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
280 GCGAGGAGCTAGAGTGGC...TGCATAACCTCATCTCTCGGAGGCG 326  
50 PheIleLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProAr 66  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
327 TAGCTCTCCAGCAATATGCGAGCGGTAACTCAACCTCACCACTTCCCGGA 376  
66 gLysAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerG 83  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
377 GAACGGCACATTGCGTGGTGAACATTGCCAGCTGAGCCAGGATGACTCCG 426  
83 lyValTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGln 99  
|| ||| ||||| |||||:|||||:|||||:||||| :|||  
427 GCGCTACAAAGTGTGGCTGGGCATCAATAGC...CGAGGCTGTGCTTT 473  
100 LysValThrLeuAsnVal 105  
|||:|||||:|||||  
474 GATGTCAGCTGGAGTGC 491





OM of: US-09-135-238B-2\_COPY\_18\_253 to: EST:\* out\_format : pfs  
Date: Sep 12, 2000 6:04 AM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:  
-MODEL=frame\_p2n.model -DEV=xlp  
-Q/cnq2.1/USPTO\_spool/SUS09135238/runat\_24082000\_091323\_17869/app\_query.fasta\_1.1282  
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -OGAPOP=4.500  
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0  
-MAXLEN=1000000 -USER=US09135238 @CN1\_1\_2178 -NCPU=6 -ICPU=3  
-LONGLOG -NO\_XLPXY -WAIT -THREADS=1

Search information block:  
Query: US-09-135-238B-2\_COPY\_18\_253  
Query length: 236  
Database: EST:\*  
Database sequences: 5247842  
Database length: -2090053206  
Search time (sec): 1563.530000

Sequence	Strd Orig	2Score	EScore	Len	Documentation
gb_est42:AW402953	+ 544.00	919.66	2.1e-42	312	! AW402953 UI-HF-BK0-abd-c-06-0-0
gb_est45:AW655210	+ 451.00	756.39	2.7e-33	518	! AW655210 105863 MARC 1BOV Bos t
gb_est42:AW444691	+ 444.50	746.13	9.9e-33	482	! AW444691 UI-H-BI3-ajw-g-12-0-UI
gb_est41:AW327067	+ 415.00	699.20	4.1e-30	359	! AW327067 20516 MARC 1BOV Bos t
gb_est45:AW655218	+ 384.00	643.17	5.4e-27	494	! AW655218 105873 MARC 1BOV Bos t
gb_est8:AA521933	+ 364.50	607.52	5.2e-25	627	! AA521933 vhl61e2.r1 Soares_mamm
gb_est12:AA830944	+ 305.00	509.27	1.5e-19	485	! AA830944 oc58b10.s1 NCI_CGAP_GC
gb_est12:AA491123	+ 299.00	509.72	1.5e-19	179	! AA491123 ac46g04.r1 NCI_CGAP_GC
gb_est40:AW291041	+ 292.50	486.49	2.9e-18	561	! AW291041 UI-H-BI2-agj-h-06-0-UI
gb_est25:AI1760373	+ 242.00	398.24	2.4e-13	711	! AI1760373 wh87d11.x1 NCI_CGAP_C
gb_est4:AA289667	+ 220.50	367.47	1.2e-11	417	! AA289667 vb20c07.r1 Soares_mous
gb_est8:AA509857	+ 209.50	356.67	4.9e-11	199	! AA509857 vhl5a10.r1 Soares_mamm
gb_est36:AA012833	+ 180.00	301.29	5.9e-08	327	! AA012833 uo3d11.y1 Sugano_mous
gb_est36:AA0174968	+ 157.50	259.46	1.3e-05	460	! AA174968 ms78h03.r1 Soares_mous
gb_est3:AA177660	+ 145.50	246.96	6.3e-05	220	! AI021484 ub08h03.r1 Soares_mamm
gb_est15:AI021484	+ 145.00	246.31	6.9e-05	216	! AA177660 mt20d11.r1 Soares_mous
gb_est4:AW603666	+ 127.00	204.60	0.0144	614	! AW603666 CM0-CN0039-280100-168
gb_gss15:AO33800	+ 121.00	196.99	0.0403	501	! AO33800 RPCI-23-248K2.TV RPCI-
gb_est48:FI14851	+ 120.50	201.00	0.0229	306	! FI14851 SSC24F11 Porcine small i
gb_est44:AW603303	+ 120.50	193.57	0.0593	614	! AW603303 RCI-CN0012-170100-011
gb_est49:H84070	+ 117.00	191.41	0.0783	431	! H84070 y94c09.r1 Soares retina
gb_est43:AW484927	+ 107.50	175.71	0.5863	414	! AW484927 63153 MARC 3BOV Bos t
gb_est12:AA839051	+ 105.50	170.54	1.14	489	! AA839051 v47a02.r1 Soares_mamm
gb_est49:H84076	+ 105.00	171.54	1.00	411	! H84076 y94c12.r1 Soares retina
gb_est17:AI177514	+ 102.50	184.73	2.40	523	! AI177514 EST221146 Normalized r
gb_est47:C88490	+ 102.50	161.47	3.64	710	! C88490 C88490 Mouse early blast
gb_est39:AW176541	+ 101.50	168.34	1.51	318	! AW176541 MR0-CT0062-200899-002
gb_est40:AW258498	+ 101.50	163.03	2.98	523	! AW258498 uq35h10.y1 NCI_CGAP_Ma
gb_est3:AA212711	+ 100.50	162.34	3.26	476	! AA212711 mw81f05.r1 Soares_mous
gb_gss8:AO349923	+ 100.50	162.03	3.39	490	! AO349923 RPCI-11-383F15.TV RPCI
gb_gss4:AO307015	+ 100.00	153.53	10.08	1003	! AO307015 HS-2046-B1-H02-77 CIT
gb_est44:AW603280	+ 99.00	157.94	5.72	566	! AW603280 RCI-CN0012-170100-011
gb_est44:AW603762	+ 96.00	155.02	8.33	462	! AW603762 QV0-ST0247-200100-092
gb_est46:AW710271	+ 96.00	154.48	8.93	486	! AW710271 e2e08ne.r1 Neurospora
gb_gss11:AO751079	+ 95.50	142.87	39.58	1332	! AO751079 HS-5576-B1-G04.SP6.RH
gb_est42:AW407899	+ 95.00	142.94	8.41	397	! AW407899 UI-HF-BM0-adk-f-09-0-U
gb_est15:AI027770	+ 94.50	146.10	26.13	839	! AI027770 ov99e09.x1 Soares_test
gb_est44:AW402453	+ 93.50	149.70	16.48	511	! AW402453 UI-HF-BK0-aan-f-12-0-U
gb_est24:AI746782	+ 93.00	152.44	11.59	365	! AI746782 ul07d03.y1 Sugano_mous
gb_est44:AW604875	+ 93.00	151.65	12.83	393	! AW604875 QV1-CT0366-280100-055
gb_est46:AW726179	+ 92.50	141.70	45.96	922	! AW726179 GA_Ea0020006 Gossypiu

gb\_est17:AI153762 + 92.00 147.98 20.55 473 ! AI153762 ud48g01.r1 Soares\_m  
gb\_est24:AI685262 - 92.00 145.93 26.72 573 ! AI685262 wa75e11.x1 Soares\_N  
gb\_est5:AA306437 + 91.50 146.52 24.79 501 ! AA306437 EST177393 Jurkat T-  
gb\_est54:W42194 - 91.50 138.06 73.30 1106 ! W42194 mc69a10.r1 Soares mo  
seq\_name: gb\_est42:AW402953  
seq\_documentation\_block: 312 bp mRNA EST 16-FEB-2000  
LOCUS AW402953  
DEFINITION UI-HF-BK0-abd-c-06-0-UI.r1 NIH\_MGC\_36 Homo sapiens CDNA clone  
IMAGE:3055859 5', mRNA sequence.  
ACCESSION AW402953  
VERSION AW402953.1 GI:6921743  
KEYWORDS EST..  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 312)  
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT On May 18, 1998 this sequence version replaced gi:3138550.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Eco RI site shown at the beginning of the sequence.  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
CDNA Library Preparation: M.B. Soares Lab  
CDNA Library Arrayed by: M.B. Soares Lab  
DNA Sequencing by: M.B. Soares Lab  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
www-bio.lnl.gov/bbrp/image/image.html  
Seq primer: M13 Forward.

FEATURES  
Location/Qualifiers  
1..312  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3055859"  
/clone\_lib="NIH\_MGC\_36"  
/tissue\_type="lymph"  
/cell\_type="germinal center B cells"  
/cell\_lines="MGC85"  
/lab\_host="DH10B (LRI)"  
/note="vector: p7T3-Pac; Site\_1: NotI; Site\_2: Eco RI;  
Constructed from size fractionated cytoplasmic mRNA  
(0.5-1.5kb). Directionally cloned. Cells provided by Louis  
M. Staudt, Ph.D. Library preparation by Maria de Fatima  
Bonaldo, Ph.D. and M. Bento Soares, Ph.D."

BASE COUNT 97 a 74 c 87 g 54 t  
ORIGIN

alignment\_scores:  
Quality: 544.00 Length: 101  
Ratio: 5.386 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-135-238B-2\_COPY\_18\_253 x AW402953 ..  
Align seg 1/1 to: AW402953 from: 1 to: 312

24 GLUMethHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyth 40  
|||||  
10 GAAATGCATGTGAGGATATATCTGTGCGGAGATGCTGGATCTGGAAC 59  
40 rCysGlyThrValValSerThrThrAsnPheIleLysAlaGluTyrLysG 57  
|||||  
60 ATGTGGTACCTGGTGTCCACCACTTCATCAAGGAGAGATCAAGG 109  
57 lYArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGlu 73

110 GCCGAGTTACTCTGAACAATACCCACGCAAGATCTGTCTAGTGGAG 159  
74 ValThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaG1 90  
160 GTAACACAGCTGACAGAAAGTGACAGCGAGTCTATGCTCGGAGCGGG 209  
90 yMetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnValHis 107  
210 CATGAACACAGACCGGGAAGACCCAGAAAGTCACCCCTGAATGTCCACA 259  
107 erGlyTyrGluProSerTrpGluGlnProMetProGluThrProLys 123  
260 GTGAATAGACCATCATCGGGAAGAGACCAATGCTGAGAGATCCAAA 309  
124 Ttp 124  
111  
310 TGG 312  
seq\_name: gb\_est45:AW655210  
seq\_documentation\_block:  
LOCUS AW655210 518 bp mRNA EST 05-APR-2000  
DEFINITION 105863 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.  
ACCESSION AW655210  
VERSION AW655210.1 GI:7421036  
KEYWORDS EST.  
SOURCE Bos taurus.  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
1 (bases 1 to 518)  
Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,  
Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A. and  
Keele,J.W.  
Design and use of four pooled tissue normalized cDNA libraries for  
EST discovery in cattle  
Unpublished (2000)  
On Mar 8, 1999 this sequence version replaced gi:4388119.  
Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smithemail.marc.usda.gov  
Single pass sequencing. Bases called and alt\_trimmed with phred  
v0.980904.e. Vector identified by cross\_match with the -minscore 18  
and -mismatch 12 options.  
PCR Primers  
FORWARD: AGGAACACAGCTATGACCAT  
BACKWARD: GTTATCCAGTCACGACG  
Plate: 86 row: K column: 10  
Seq primer: ATTTAGTGACACTATAG.  
FEATURES  
source  
1..518  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_lib="MARC 1BOV"  
/tissue\_type="pooled"  
/lab\_host="DH10b"  
/note="Vector: pCMV SPORT6; site.1: XbaI; site.2: XhoI;  
Library made from pooled tissue from lymph node, ovary,  
fat, hypothalamus, and pituitary."  
124 a 133 c 157 g 104 t  
BASE COUNT 124 a 133 c 157 g 104 t  
ORIGIN  
alignment\_scores:  
Quality: 451.00 Length: 119  
Ratio: 4.295 Gaps: 0  
Percent Similarity: 88.235 Percent Identity: 67.227

alignment\_block:  
US-09-135-238b-2\_copy\_18\_253 x AW655210 ...  
Align seg 1/1 to: AW655210 from: 1 to: 518  
1 ArgIleLeuProGluValLysValGluGlyGluLeuGlyGlySerValTh 17  
161 AAGTCTCTCCAGAAAGTAAGATGAAGGAATGCTGGGAGGATCCATTAC 210  
17 rIleLysCysProLeuProGluMetHisValArgIleTyrIleuCysArg 34  
211 CATCGATGCCCACTCTCTGAACGCATGTGAGGATATCTGTGCCGGA 260  
34 LuMetalGlySerGlyThrCysGlyThrValValSerThrThrAsnPh 50  
261 CCATTGACGAGTCTGGAAGATGCCACCGTGTGTCTCCAGCAACAATAC 310  
51 IleLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProArg 67  
311 GTCAGAGAGGAATTCACGCCGAGTCACGCTGGAGCAGTGTCCGGACAG 360  
67 sAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerGly 84  
361 GAATCTGTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 410  
84 alTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGln 100  
411 TCTATGCTGTGGGTGGCGGGNACACAGACCCGAGGAGCAGACCCAG 460  
101 ValThrLeuAsnValHisSerGluTyrGluProSerTrpGluGluGln 117  
461 ATCACCCTGACTGTTTCACAGCGTTTACGAGCCATCTCTGGGAAGA 510  
117 oMetPro 119  
511 GATGCCT 517  
seq\_name: gb\_est42:AW444691  
seq\_documentation\_block:  
LOCUS AW444691 482 bp mRNA EST 17-FEB-2000  
DEFINITION UI-H-BI3-ajw-g-12-0-UI-s1 NCI CGAP\_Sub5 Homo sapiens cDNA clone  
IMAGE:2733502 3', mRNA sequence.  
ACCESSION AW444691  
VERSION AW444691.1 GI:6986453  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 482)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
The sequence contained an oligo-dT track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. cDNA Library preparation: M.B. Soares Lab Clone distribution:  
NCI-CGAP clone distribution information can be found through the  
I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html) The following repetitive  
elements were found in this cDNA sequence: 1-30,  
>POLY\_A\$Simple\_repeat  
Seq primer: M13 Forward  
POLYA=Yes,  
FEATURES  
source  
1..482  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

```
/clone="IMAGE:2733502"
/clone_lib="NCI_CGAP_Sub5"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pT713D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; NCI_CGAP_Sub5
is a subtracted library derived from NCI_CGAP_Sub4. The
NCI_CGAP_Sub5 library had 3 million recombinants. A
single-stranded DNA preparation of NCI_CGAP_Sub4 was used
as a tracer in a subtractive hybridization with a driver
comprising: the IMAGE pool (NCI_CGAP_Kid3 pool 1 LLAM
3334-3337, 3682-3683, 3798-3803 (IMAGE Clonoids
1322376-1323911, 1456008-1456775, 1500552-1502855);
NCI_CGAP_Kid5 pool 1 LLAM 3338-3342, 3723-3725, 3776-3778
(IMAGE Clonoids 1323912-1325831, 1471368-1472903
1492104-1493255); NCI_CGAP_Lu5 pool 1 LLAM 3575-3582,
3851-3854 (IMAGE Clonoids
1414920-1417991, 1520904-1522439); NCI_CGAP_GC4 pool 1 LLAM
3164-3167, 3716-3720, 3733-3735 (IMAGE Clonoids
1257096-1258631, 1469064-1470983, 1475592-1476743);
NCI_CGAP_Pr22 pool 1 LLAM 2457-2459, 2758-2759, 3062-3068
(IMAGE Clonoids 985608-986759, 1101192-1101959,
1217928-1220615); NCI_CGAP_Co10 pool 1 LLAM 2644-2653,
2871-2872 (IMAGE Clonoids
1057416-1061255, 1144584-1145351). (10% of the driver
population), plus a pool of 3,840 arrayed clones from
NCI_CGAP_Sub1 (IMAGE Clonoids 2708616-2710535) and
NCI_CGAP_Sub2 (IMAGE Clonoids 2710536-2712455) (10% of
the driver population), plus a pool of 11,136 clones from
NCI_CGAP_Sub3 (IMAGE Clonoids 2712456-2723591) (10% of the
driver population), plus a pool of 5,472 clones from
NCI_CGAP_Sub4 (IMAGE Clonoids 2723592-2728969) (70% of the
driver population). Subtraction was performed as
previously described [Bonaldo, Lennon & Soares (1996):
Normalization and Subtraction: Two Approaches To
Facilitate Gene Discovery. Genome Research 6, 791-806. ;
TAG_LIB=NCI_CGAP_Kid5; TAG_TISSUE=Kidney; TAG_SEQ=ATTC"
BASE COUNT      103 a 107 c 111 g 161 t
ORIGIN

alignment_scores:
  Quality: 444.50      Length: 117
  Ratio: 4.401        Gaps: 3
  Percent Similarity: 86.325  Percent Identity: 80.342

alignment_block:
US-09-135-238B-2_COPY_18_253 x AW444691/rev  ..
Align seg 1/1 to reverse of: AW444691 from: 1 to: 482

24  GluMetHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyTh 40
|||||
481  GAAATGCATGTGAGGATATCTGTCCGGAGATGGCTGGATCTGGAAC 432
|||||
40  rCysGlyThrValValSerThrThrAsnPheIleLysAlaGluTyrLysG 57
|||||
431  ATGTGTACGTGGTATCCACCACCACTTCATCAAGGCAGATACAAAG 382
|||||
57  lyArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGlu 73
|||||
381  GCCGAGTTACTCTGAAGCAATACCCACGAAGAATCTGTTCTAGTGGAG 332
|||||
74  ValThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGl 90
|||||
331  GTAACACAGCTGACAAAGTGTACAGCGGAGTCTATGCTCGGAGCGGG 282
|||||
90  yMetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnValHis 107
|||||
281  CATGAACACAGACCGGGGAAGAACCCAGAAAGTCACCCCTGAATGTCCA 232
|||||
107  er...GluTyrGluProSerTrpGluGluGlnProMetProGluThrPro 122
|| :|:|:|
231  GTGGTAGGTTCCCTCCGCTGATTGGAGGCTCAG.....ACTACCCAG 191
```

```
123  LysTrpPheHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSerSe 139
||| ||||| ||| :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
190  AAAATGTTTCACTTGGGAAGATTG...GAACAGCCACTCTAACTGTAACAG 144
139 r 139
|
143 T 143

seq_name: gb_est41:AW327067

seq_documentation_block:
LOCUS      AW327067      359 bp      mRNA      EST      27-JAN-2000
DEFINITION 20516 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION  AW327067
VERSION    AW327067.1 GI:6762988
KEYWORDS   EST.
SOURCE     Bos taurus.
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE  1 (bases 1 to 359)
AUTHORS   Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A. and
Keefe,J.W.
TITLE     Design and use of four pooled tissue normalized cDNA libraries for
EST discovery in cattle
JOURNAL   Unpublished (2000)
COMMENT   On Jan 6, 2000 this sequence version replaced gi:6676592.
Contact: Smith Trl
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4360
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 20
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCGACGTCACGACG
Plate: 13 row: K column: 5
Seq primer: ATTAGGTGACACTATAG.
Location/Qualifiers
  1..359
  /organism="Bos taurus"
  /db_xref="taxon:9913"
  /clone_lib="MARC 1BOV"
  /tissue_type="pooled"
  /lab_host="DH10B"
  /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from lymph node, ovary,
fat, hypothalamus, and pituitary."
BASE COUNT      90 a 111 c 87 g 71 t
ORIGIN

alignment_scores:
  Quality: 415.00      Length: 119
  Ratio: 4.192        Gaps: 1
  Percent Similarity: 83.193  Percent Identity: 64.706

alignment_block:
US-09-135-238B-2_COPY_18_253 x AW327067  ..
Align seg 1/1 to: AW327067 from: 1 to: 359

59  ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75
|||||
2  GTCACGCTGGACAGTGTCCGGACAGGAATCTGTTCTCTGTGTGTATGAC 51
|||||
75  rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 92
```

```

|||||
52 AGAGCTGACCAAGAAATGACAGCGGGATCTATGCTGTGGGGTGGCGGGA 101
|||||
92 snThrAspArgGlyysThrGlnLysValThrLeuAnValHisSerGlu 108
|||||
102 ACACAGACGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 151
|||||
109 TyGluProSerTrpGluGluGlnProMetProThrThrProLysTrpPh 125
|||||
152 TAGAGCCATCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 201
|||||
125 e.....HisLeuProTyThrLeuPheGlnMetProAlaIat 136
|||||
202 TAATAGATTCTACAAATGCACCTGCTCCCTTGGTTCAGATGCCTGCAC 251
|||||
136 yrlaSerSerLysPheValThrArgValThrThrProAlaGlnArg 152
|||||
252 ATGCCAGTTCTTTAAATTCACGTCAAGTAACACACACAGCTCAAGG 301
|||||
153 GlyLysValProValHisHisSerSerProThrThrGlnIleThrH1 169
|||||
302 ATCAAGTCTCTCGAGCACACAGGCTCCCCCAACCCCTCAGTCACCCA 351
|||||
169 sArgPro 171
|||||
352 CCGCCCC 358

seq_name: gb_est45:AW655218

seq_documentation_block: 494 bp mRNA EST 05-APR-2000
LOCUS AW655218
DEFINITION 105873 MARC 1BOV Bos taurus CDNA 5', mRNA sequence.
ACCESSION AW655218
VERSION AW655218.1 GI:7421044
KEYWORDS EST.
ORGANISM Bos taurus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 494)
Smith,R.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A. and
Keele,J.W.
Design and use of four pooled tissue normalized cDNA libraries for
EST discovery in cattle
Unpublished (2000)
Contact: Smith TPL
On Mar 8, 1999 this sequence version replaced gi:4388128.
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt.trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACACGCTATGACCAT
BACKWARD: GTTTCCAGTCACGACG
Plate: 86 row: M column: 8
Seq primer: ATTTAGGTGACACTAG.
FEATURES
Source
1..494
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 1BOV"
/tissue_type="pooled"
/lab_host="Dhl08"
/note="Vector: pCMV SPORT6; Site.1: XbaI; Site.2: XhoI;
Library made from pooled tissue from lymph node, ovary,
fat, hypothalamus, and pituitary."
BASE COUNT 118 a 125 c 150 g 101 t

```

```

ORIGIN
alignment_scores:
  Quality: 384.00 Length: 111
  Ratio: 4.042 Gaps: 0
Percent Similarity: 85.586 Percent Identity: 63.063

alignment_block:
US-09-135-238B-2_COPY_18_253 x AW655218 ..
Align seg 1/1 to: AW655218 from: 1 to: 494
1 ArgIleLeuProGluValLysValGluGlyGluLeuGlyGlySerValTh 17
162 AAGTGCTTCCAGAGTAAGATGAAGAAATGCTGGGAGGATCCATTC 211
17 rIleLysCysProLeuProGluMetHisValArgIleTyrLeuCysArg 34
212 CATCGAGTGCCCACTTCCTGAAACGCATGTGAGGATATATCTGTGCGGA 261
34 LuMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsnPhe 50
262 CCATGTGAGAGCTGGAAGATGCACCACTGTGTCCAGCAACAATAC 311
51 IleLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProArgLy 67
312 GTAAGAGAGGAATTCAAGCAGGAGGACGCTGGAGCAGTGTCCGGACAG 361
67 sAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerGlyV 84
362 GAATCTGTCTTGTGTGATGACGAGCTGACCAAGATGACAGCGGGA 411
84 alTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLys 100
412 TCTATGCTGTGGGTGGGGGAACACAGACGAGGAGCAAGACCCATCAG 461
101 ValThrLeuAsnValHisSerGluTyrGluPro 111
462 ATCACCCTGACTGTTACAGCGTTTACTAGCCA 494

seq_name: gb_est8:AA521993

seq_documentation_block: 627 bp mRNA EST 17-JUL-1997
LOCUS AA521993
DEFINITION wh61e12.r1 Soaresmammary.gland_NbMMG Mus musculus CDNA clone
IMAGE:891502 5', mRNA sequence.
ACCESSION AA521993
VERSION AA521993.1 GI:2262738
KEYWORDS EST.
ORGANISM house mouse.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 627)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
On Jan 14, 1998 this sequence version replaced gi:1797443.
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MG1:519462

```



162 SerProThrThrGlnIleThrHisArgProArgValSerArgAlaSerSe 178  
 |||||  
 485 TCCGCCACACCCAAATACACCGCCCTCGAGTGTCCAGAGCATCTTC 436  
 |||||  
 178 rValAlaGlyAspLysProArgThrPheLeuProSerThrThrAlaSerL 195  
 |||||  
 435 AGTAGCAGGTGACAAAGCCCGAACCTTCTGCTCCATCCACTACAGCCTCAA 386  
 |||||  
 195 ystIleSerAlaLeuGluGlyLeuLeuLysProGlnThrProSerTyrAsn 211  
 |||||  
 385 AAATCTCAGCTCTGGAGGGCTGCTCAAGCCCCAGAGCCGACGCTACAAAC 336  
 |||||  
 212 HisHisThrArgLeuHisArgGlnArg 220  
 |||||  
 335 CACCACACGAGCTGCACAGGAGGAGG 309

seq\_name: gb\_est8:AA491123

seq\_documentation\_block: 179 bp mRNA EST 18-AUG-1997  
 LOCUS AA491123  
 DEFINITION aa46q04.r1 NCI\_CGAP\_GCB1 Homo sapiens cDNA clone IMAGE:824022 5',  
 mRNA sequence.  
 ACCESSION AA491123  
 VERSION AA491123.1 GI:2220296  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 179)

REFERENCE  
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index

JOURNAL  
 COMMENT Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,  
 Ph.D., Gerald Marti, M.D.  
 CGAP Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
 Bonaldo, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

Seq primer: -28m13 rev1 Et from Amersham  
 High quality sequence stop: 116.

## FEATURES

source

Location/Qualifiers  
 1..179  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:824022"  
 /clone\_lib="NCI\_CGAP\_GCB1"  
 /tissue\_type="germinal center B cell"  
 /lab\_host="DH10B"  
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
 polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA  
 was prepared from human tonsillar cells enriched for  
 germinal center B cells by flow sorting (CD20+, IgD-),  
 provided by Dr. Louis M. Staudt (NCI), Dr. David Allman  
 (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was  
 primed with a Not I - oligo(dT) primer  
 [5'-TGTTACCAATCTGAAGTGGAGCGCGCTCAATTTTTTTTTTTT-3'  
 ]. Double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pT7T3 vector. Library  
 went through one round of normalization, and was  
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT  
 ORIGIN

alignment\_scores:  
 Quality: 299.00 Length: 59  
 Ratio: 5.068 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-09-135-238B-2\_COPY\_18\_253 x AA491123 ..

Align seg 1/1 to: AA491123 from: 1 to: 179

146 ValThrThrProAlaGlnArgLysValProProValHisHisSerSe 162  
 |||||  
 3 GTTACCACACGAGCTCAAAGGGGCAAGTCCCTCCAGTTCACCACCTCCTC 52  
 |||||  
 162 rProThrThrGlnIleThrHisArgProArgValSerArgAlaSerSe 179  
 |||||  
 53 CCCACACCCCAAAATCACCCACGAGCTCGAGTGTCCAGAGCATCTTCAG 102  
 |||||  
 179 alAlaGlyAspLysProArgThrPheLeuProSerThrThrAlaSerLys 195  
 |||||  
 103 TAGCAGGTGACAAAGCCCGAACCTTCTGCTCCATCCACTACAGCCTCAAAA 152  
 |||||  
 196 IleSerAlaLeuGluGlyLeuLeuLys 204  
 |||||  
 153 ATCTCAGCTCTGGAGGGGCTGCTCAAG 179

seq\_name: gb\_est40:AW291041

seq\_documentation\_block: 561 bp mRNA EST 16-JAN-2000  
 LOCUS AW291041  
 DEFINITION UT-H-B12-agi-h-06-0-OI.s1 NCI\_CGAP\_Sub4 Homo sapiens cDNA clone  
 IMAGE:2724707 3', mRNA sequence.

ACCESSION AW291041

VERSION AW291041.1 GI:6697677

KEYWORDS EST.

SOURCE human.

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS 1 (bases 1 to 561)

TITLE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

Journal Tumor Gene Index

Unpublished (1997)

COMMENT On May 16, 1998 this sequence version replaced gi:3137211.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

The sequence contained an oligo-dT track that was present in the  
 oligonucleotide that was used to prime the synthesis of first  
 strand cDNA and therefore this may represent a bonafide poly A  
 tail. cDNA library Preparation: M.B. Soares Lab Clone distribution:  
 NCI-CGAP clone distribution information can be found through the  
 I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

Seq primer: M13 Forward

POLYA-Yes.

Location/Qualifiers

1..561

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:2724707"

/clone\_lib="NCI\_CGAP\_Sub4"

/lab\_host="NCI\_CGAP\_Sub4"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified

polylinker. Site 1: Not I; Site 2: Eco RI; The

NCI\_CGAP\_Sub4 library is a subtracted library derived from  
 the NCI\_CGAP\_Sub2 library which is a subtracted library  
 derived from the NCI\_CGAP\_Sub1 library, which is a  
 subtracted library derived from BI. BI constitutes a  
 mixture of 21 normalized or subtracted  
 libraries: NCI\_CGAP\_C04, NCI\_CGAP\_P22, NCI\_CGAP\_Pr28,  
 NCI\_CGAP\_C010, NCI\_CGAP\_C016, NCI\_CGAP\_K105,

EST.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 711)  
1 (bases 1 to 711)  
NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On Oct 30, 1998 this sequence replaced gi:3812483.

source

```

BASE COUNT
ORIGIN
alignment_s
Percent Sili
alignment_b
US-09-135-
Align seg
173 Va
||
711 GT

```

```

206 ln
    ||
611 AG

222 uA
    ||
561 GG

seq_name: g

```

seq\_name: gb\_est4:AA289667

[illegible]







206 InThrProSerTyrAsnHisHisThrArgLeuHisArgGlnArgAlaLeu 222  
| | | | | : : : | | | | | | | | | | | | | | | : : : | | | | | : : :  
98 TAGAGGCCAGCTACAGCCACCACACCCAGACTTCATGAGCAAGGACACGC 147  
223 AspTyrGlySerGlnSerGlyArgGluGlyGlnGlyPheHis 236  
: : : | | : : : | | | | | | | : : : | | | | | : : : | | |  
148 CACCATGGCCACACTATGGGAGAGAGAGACCGAGGGGCTTCAC 189





```

181 GCGAGCTGTTTCATACGCTGCTACTACCCAGACACCTCTGTCAACCG 230
25 thisValArgileTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 42
231 GCACACCCGGAATACTGTGGCGACAAGGAGCC...AGCGCATGTGCA 277
42 lyThrValValSerThrThrAsnPhelLeuLysAlaGluTyrLysGlyArg 58
278 CAACGTCATCTCTCAATGGCTACCTCTCCAAGGAGATTACAGGAGA 327
59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75
328 GCCAACCTCATCACTCCCGACAGAACACACATTGTGATTAACATGA 377
75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 92
378 GCAGCTCACCCAGGACGACACTGGGAGCTCAAGTGTGGCTGGGT...A 424
92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 108
425 CCAGTAACCGAGCGCTGCTCCTCGATGTGAGCTGGAGTC..... 465
109 TyrGluProSerTyrGluGluGlnProMetProGluThrProLysTyrPh 125
466 .....AGCCAGGTTCTGAGTGGCGAGTGACAC 494
125 eHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSerSerLysP 142
495 CCACGTC.....TACAAAGGACATAGCCA 520
142 heValThrArgValThrProAlaGlnArgGlyLysValPro 156
521 GAATGTGACCATGAATGCCCTTTCAAAGGGAGAATGTGCC 564

seq_name: /cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-434-000A-9
seq_documentation_block:
; Sequence 9, Application US/08434000A
; Patent No. 6046037
; GENERAL INFORMATION:
; APPLICANT: ANDREW C. HIATT, JULIAN
; APPLICANT: K.-C. MA, THOMAS LEHNER
; TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
; TITLE OF INVENTION: PROTEINS IN PLANTS AND THEIR USES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/434,000A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below:
; APPLICATION NUMBER: 08/367,395
; FILING DATE: 12/30/94
; ATTORNEY/AGENT INFORMATION:
; NAME: Guise, Jeffrey W.
; REGISTRATION NUMBER: 34,613
; REFERENCE/DOCKET NUMBER: 212/127
;

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 552-8400
; TELEFAX: (619) 552-0159
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3269 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 74....2383
; US-08-434-000A-9

alignment_scores:
Quality: 185.50 Length: 164
Ratio: 1.855 Gaps: 6
Percent Similarity: 60.976 Percent Identity: 30.488

alignment_block:
US-09-135-238B-2_copy_18_253 x US-08-434-000A-9 ..
Align seg 1/1 to: US-08-434-000A-9 from: 1 to: 3269

13 GlyGlySerValThrIleLeuLysCysProLeuProGlu.....Me 25
170 GGTAACTCGGTCTCCATACGCTACTACCCAGACACCTCTGTCAACCG 219
25 thisValArgileTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 42
220 GCACACCCGGAATACTGTGGCGACAAGGAGCC...AACGCTACTGCG 266
42 lyThrValValSerThrThrAsnPhelLeuLysAlaGluTyrLysGlyArg 58
267 CAACCTCATCTCTTCATATGGCTACTCTCGAAGGAGTATTCAGGAGA 316
59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75
317 GCGACGCTCATCACTCCCGACAGATAGCACATTGTGATTAACATGC 366
75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 92
367 ACATCTCACCCAGGAGGACACTGGGAGCTACAAAGTGTGCTGGGT...A 413
92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 108
414 CCACCTAACCGAGGCGCTGTTTTCGATGTGACGCTGGAGTC..... 454
109 TyrGluProSerTyrGluGlnProMetProGluThrProLysTyrPh 125
455 .....AGCCAGGTTCTGAGTGGCGAGTC..... 483
125 eHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSerSerLysP 142
484 CCATGTC.....TACAAAGG 500
142 heValThrArgValThrProAlaGlnArgGlyLysValProVal 158
501 ACATAGGCGAAGTGTGACCATCGAATGCCGTTTCAAGAGGGGAATGCT 550
159 HisHisSer.SerProThrThrGlnIleThrHisArgPro 171
551 CATAGCAAGAAATCCCTGTGTAGGAAGAGAGAGAGGCGCT 590

seq_name: /cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-434-000A-3
seq_documentation_block:
; Sequence 3, Application US/08434000A
; Patent No. 6046037
```









```
25 thisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 42
   ||| ||| :||| |||
288 GCACAGCGGAAGTTCTGGTCGGGAA...GAGGAGAGCGCGCGTGC 334
   ||| ||| :||| |||
42 lYthrValValSerThrThrAsnPheIleLysAlaGluTyrIlySgIArg 58
   ||| ||| :||| |||
335 TGACGCTTGCCTCG...ACCGGTACAGTCCCGAGGAATACTCCGGGAGA 381
   ||| ||| :||| |||
59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75
   ||| ||| :||| |||
382 GCAAGCTCACCGACTCCCTCGATAAAGGGGAGTTGTGGTGA 431
   ||| ||| :||| |||
75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMet 92
   ||| ||| :||| |||
432 CCAACTCACCAAGAGACTCAGGAGCTACAAGTGTGGCGGAGTCA 481
   ||| ||| :||| |||
92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 108
   ||| ||| :||| |||
482 AC...GGCGGTGGCGTGGACTTCGGTGTCAACGTGCTGTCAGCCAGAAG 528
   ||| ||| :||| |||
109 TyrGluPro 111
   ||| ||| :||| |||
529 CCAGAGCGCT 537
```

seq\_name: /cgn2\_6/ptodata/1/ina/5A\_comb.seq:US-08-286-740-3

```
seq documentation block:
; Sequence 3: Application US/08286740
; Patent No. 5561053
; GENERAL INFORMATION:
; APPLICANT: Crowley, Craig W.
; TITLE OF INVENTION: METHOD FOR SELECTING HIGH-EXPRESSING
; TITLE OF INVENTION: HOST CELLS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/286,740
; FILING DATE: 05-AUG-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee, Wendy M.
; REGISTRATION NUMBER:
; REFERENCE/DOCKET NUMBER: 798
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1994
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6557 bases
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-286-740-3
```

alignment\_scores:
Quality: 102.00
Ratio: 0.843
Length: 252
Gaps: 15

```
Percent Similarity: 48.016 Percent Identity: 24.603
alignment_block:
US-09-135-238b-2_copy_18_253 x US-08-286-740-3 ..
Align seg 1/1 to: US-08-286-740-3 from: 1 to: 6557
13 GlyGlySerValThrIleLysCysProLeuProGluMethHisValArg11 29
   ||| ||| :||| |||
1405 GGGGGCTCACTCGGTTGCTCTGCGAGTTCTGGCTACTCCATCACCTC 1454
   ||| ||| :||| |||
29 e.....TyrLeuCysArgGluMetAlaGlySerGly..... 39
   ||| ||| :||| |||
1455 CGGATATAGCTGGAACCTGGATCGCTAGCGCCCGGGTAAGGCGCTGGAAT 1504
   ||| ||| :||| |||
40 .....ThrCysGlyThrValValSerThrThrAsnPheIleLysAlaGlu 54
   ||| ||| :||| |||
1505 GGGTTCATCGATTACGTATGCGGATCGACTACTAT...AACCTAGC 1551
   ||| ||| :||| |||
55 TyrIlySgIArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLe 71
   ||| ||| :||| |||
1552 GTCAGGCGCGTATCACTATAAGTCGCGAGCATTCACAAAACACATCTCA 1601
   ||| ||| :||| |||
71 uValGluValThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysG 88
   ||| ||| :||| |||
1602 CTGCGAGATGAACAGCGCTGGCTGAGGACACTGCGGCTCTATTATTGTG 1651
   ||| ||| :||| |||
88 lYalaglyMetAsn..ThrAspArgGlyLysThrGlnLysValThrLeuAs 104
   ||| ||| :||| |||
1652 CTCGAGGAGCAGCACTATTTCGGCGCTGCGCACTTCGCGCTGGGTCAA 1701
   ||| ||| :||| |||
104 nValHisSerGluTyrGluProSerTrpGluGluInProMetPro.... 119
   ||| ||| :||| |||
1702 G.....GAACCC...TGGTACCGCTCTCTCGGCCCTCCAC 1733
   ||| ||| :||| |||
120 .....GluThrProLysTrpPhe.....HisLeuProTyr 129
   ||| ||| :||| |||
1734 CAAGGCGCCATCGGTCTTCCCGCTGGCACCCCTCTCCAGAGACACCTGTG 1783
   ||| ||| :||| |||
130 LeuPheGlnMetProAlaTyrAlaSerSerSerLysPheValThrArgVa 146
   ||| ||| :||| |||
1784 GGGGCACAGCGCCCTGGGTGCTGTCTCA.....AGGAC 1818
   ||| ||| :||| |||
146 lThrThrProAlaGln.....ArgGlyLysValProProValHisH 160
   ||| ||| :||| |||
1819 TACTTCCCGAAGCGGTGACGGTGTCTGGAACCTCAGGCGCGCTGACCAG 1868
   ||| ||| :||| |||
160 isSerSerProThrThrGlnIleThrHisArgProArgValSerArgAla 176
   ||| ||| :||| |||
1869 CGGGTGCACACTTCCGGGCTGTCTAGTCTCTCAGGACTCTACTGCC 1918
   ||| ||| :||| |||
177 SerSer.....ValAlaGlyAspLysProArgThrPh 187
   ||| ||| :||| |||
1919 TCAGCAGCGTGGTGACTGTCCCTCTAGCAGCTTGGGCACCCAGA..... 1963
   ||| ||| :||| |||
187 eLeuProSerThrThrAlaSerLysIleSerAlaLeuGluLysLeuL 204
   ||| ||| :||| |||
1964 .....CCTACATCTGCAACGTGAATCACAGC..... 1990
   ||| ||| :||| |||
204 ysProGlnThrPro.....SerTyrAsnHisHisThr 214
   ||| ||| :||| |||
1991 ..CCAGCAACACCAAGGTGGACAAGAAAGTTGAGCCCAAAATCTGTGTACA 2038
   ||| ||| :||| |||
215 ArgLeu...HisArgGlnArgAlaLeuAspTyrGlySerGlnSerGlyAr 230
   ||| ||| :||| |||
2039 AAATCTACACATGCCCGCCCGTCCCGACGACCTGGAACCTCTGCGGGGACCG 2088
   ||| ||| :||| |||
230 gGlu 231
   ||| ||| :||| |||
2089 TCAG 2092
```

seq\_name: /cgn2\_6/ptodata/1/ina/PCTUS\_comb.seq:PCT-US95-09576-3



```

: ATTORNEY/AGENT INFORMATION:
: NAME: ERNST, BARBARA G
: REGISTRATION NUMBER: 30,377
: REFERENCE/DOCKET NUMBER: 1808-118
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 783-6040
: TELEFAX: (202) 783-6031
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1617 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: both
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: FEATURE:
: NAME/KEY: sig_peptide
: LOCATION: 35..92
: FEATURE:
: NAME/KEY: mat_peptide
: LOCATION: 93..1465
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 35..1465
: US-08-378-939-9

```

```

alignment_scores:
  Quality: 100.50      Length: 254
  Ratio: 0.817        Gaps: 11
  Percent Similarity: 48.425  Percent Identity: 24.016

```

```
alignment_block:
```

```
US-09-135-238B-2_COPY_18_253 x US-08-378-939-9 ..
```

```
Align seg 1/1 to: US-08-378-939-9 from: 1 to: 1617
```

```

5  GluValLysValGluGlyLeuGlySerValThrIleLysCysPr 21
   |||||
119 GAAGTAAAG.....AAGCTGGGTCTCGGTGACGGTCTCTGCAA 159
21  OleuProglu.....MetHisValArgIleTyrLeuCysArgGluM 35
   |||||
160 GGCATCTCGAGGACCTTCAGCAACTATGCTATCATCGTGGTGGCAGG 209
35  eTalaGlySerGly.....ThrCysGly 42
   |||||
210 CCCTGGACAAAGGCTTGAGTGGAGGAGGATCATCCCTTTTGGT 259
43  ThrValValSerThrThrAsnPheIleLysAlaGluTyrLysGlyArgVa 59
   |||||
260 ACACCAACTTACTTCACAGAACTC.....CAGGGCAGAGT 294
59  lThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValThrG 76
   |||||
295 CAGGATTACCCGGCAATCCACGACAGCCACATGGAGTGACTA 344
76  lLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGly..... 90
   |||||
345 GCCTGAGATCTGAGGACAGCGCGGTGATTACTGTGCGACAGATCGCTAC 394
91  .....MetAsnThrAspArgGlyLysThr.GlnLysValThrLeuAsnV 105
   |||||
395 AGCAGGCAAAATTTGACCGGCCCGGTGGTGGTTCGACCCCTGGG 444
105  aHisSerGluTyrGluProSerTrpGluGlnProMetProGluThr 121
   |||||
445 CCAGGGACACCTGGTTCACCGTCTCTCTCAGCCTCCACCAAGGGGCCATCGG 494
122  ProLysTrpPheHisLeuProTyrLysPheGlnMetProAlaTyrAlaSe 138
   |||||
495 TCTTCCCCCTGGCACCTCTCTCCAAAGACACCTCTGGGGGCACAGCGGCC 544

```

```

138 rSerSerLysPheValThrArgValThrThrProAlaGln.....A 152
   ::::
545 CTGGCTGGCTGG...TCAGGACTACTTCCCGGACGGGTGAGGGTGTTC 591
152 rgGlyLysValProProValHisSerSerProThrThrGlnIleThr 168
   |||||
592 GTGGAACACTCAGGCGCCTCACCAGCGCGTGCACACCTTCCCGGCTGTCC 641
169 HisArgProArgValSerArgAlaSerSerValAlaGlyAspLysProAr 185
   ::::
642 TACAGTCTCTCAGGACTTACTCCTCAGCAGCGGTGGTGACCGTGCCT.. 689
185 gThrPheLeuProSerThrThrAlaSerLysLysIleSerAlaLeuGluGlyL 202
   |||||
690 .....CCAGCAGCTTGGGACCCAGACCTACATCTGCAACGTGAA 729
202 euLeuLysProGlnThrPro.....SerTyrAsnHis 212
   ::::
730 TCACAGCCCGACACACCAAGGTGGACAGAAAGTTGACGCCCAATCTT 779
213 HisThrArgLeu...HisArgGlnArgAlaLeuAspTyrGlySerGlnSe 228
   |||||
780 GTGACAAAACTCACACATGCCCGTGCACCGTGCACACCTGAACTCCTGGGG 829
228 rGlyArgGlu 231
   ::::
830 GGACCGTCAG 839

```

```
seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-935-450-10
```

```
seq_documentation_block:
```

```

: Sequence 10, Application US/08935450
: Patent No. 5977311
: GENERAL INFORMATION:
: APPLICANT: Nandabalan, Krishnan
: APPLICANT: Yang, MeiJia
: APPLICANT: Schulz, Vincent
: TITLE OF INVENTION: 53BP2 COMPLEXES
: FILE REFERENCE: 7934-054
: CURRENT APPLICATION NUMBER: US/08/935,450
: CURRENT FILING DATE: 1997-09-23
: NUMBER OF SEQ ID NOS: 15
: SOFTWARE: Patent In Ver. 2.0
: SEQ ID NO 10
: LENGTH: 914
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-08-935-450-10

```

```

alignment_scores:
  Quality: 99.00      Length: 146
  Ratio: 1.456        Gaps: 9
  Percent Similarity: 46.575  Percent Identity: 29.452

```

```
alignment_block:
```

```
US-09-135-238B-2_COPY_18_253 x US-08-935-450-10 ..
```

```
Align seg 1/1 to: US-08-935-450-10 from: 1 to: 914
```

```

116 GluProMetProGluThrProLysTrpPheHisLeu..... 127
   |||||
232 CAGCCCTGCTCGGAACCCCGGCGCCAGCACCTACATAAGAAAGACA 281
128 ....ProTyrLeu.....PheGlnMetProAlaTyrAlaSerSerS 140
   |||||
282 ACATCCCTGGCTCAAGGCCAATACCAGCACCCCGCCGTCAGCAGCTAC 331
140 erLysPheValThrArgValThr.....ProAlaGlnArg..... 152
   ::::
332 AGCCCTTCCACAGCGGAGTTACAGCCAGCCACCTACACACGAGGAGGT 381

```

```
153 .....GlyLysValProPro.ValHisHisSerSerP 163
154 ||||| :|||||
382 TACAGCCAGGTTTACAGGCCACCGCTCCACCTCCACCCACCTGC 431
163 rothrThrGln.....IleThrHisArgProArgValSerArg 175
164 ||||| :|||||
432 CTACAACTATGGAGTACCGCGGTACACCCGCGCCCTATATACCCAC 481
176 AlaSerSerValAlaGlyAspLysProArgThrPheLeuProSerThr 192
482 CGCCACCCC...CCAGCCGACAGACCTACCTCAGCCCACTATAACCCAG 528
192 rAlaSerLysIleSerAlaLeuGluGlyLeuLeuLys.....ProG 206
529 TATCAGCAGTATGCCAGCTACGAGTGAACACAGTACTCAGAACACAGGCCAG 578
206 lnThrProSerTyrAsnHisHisThrArgLeuHisArgGlnArgAlaLeu 222
579 TGGCGCCATACTACGGGAATAC..... 601
223 AspTyrGlySerGlnSerGlyArgGluGlyGlnGly 234
602 GACTACGGGAGCTACTCCGGGAACACACAGGGTGGC 637
```

seq\_name: /cgn2\_6/ptodata/1/lna/5D\_COMB.seq:US-08-887-352B-1

```
seq_documentation_block:
; Sequence 1, Application US/08887352B
; Patent No. 5994511
; GENERAL INFORMATION:
; APPLICANT: Henry B. Lowman, Leonard G. Presta, Paula M. Jardieu, John Lowe
; TITLE OF INVENTION: Improved Anti-IgE Antibodies and Method of
; TITLE OF INVENTION: Improving Polypeptides
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; FILING DATE: 03-Jul-1997
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1123
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6127 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Double
; TOPOLOGY: Circular
US-08-887-352B-1
```

```
alignment_scores:
  Quality: 98.50      Length: 220
  Ratio: 0.947       Gaps: 10
Percent Similarity: 47.273      Percent Identity: 23.636
alignment_block:
US-09-135-238B-2_COPY_18_253 x US-08-887-352B-1 ..
```

```
Align seg 1/1 to: US-08-887-352B-1 from: 1 to: 6127
13 GlyGlySerValThrIleLysCysProLeuProGluMethHisValArgII 29
1357 GGGGCTCACTCCGTTTGTCTCTGTCAGTTTCTGGCTACTCATCACTC 1406
29 e.....TyrLeuCysArgGluMetAlaGlySerGly..... 39
1407 CGGATACAGCTGGAACTGGATCCGTCAGCCCGGGTAAGGGCCCTGGAAT 1456
40 .....ThrCysGlyThrValValSerThrThrAsnPheIleLysAlaGlu 54
1457 GGGTTCATCGATTAAGTATGACGATGACGACTAATAT...AACCTAGC 1503
55 TyrLysGlyArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLe 71
1504 GTCAAGGGCCGTATCATATTAAGTCGCGCAGCATTCACAAAAACATCTCA 1553
71 uValGluValThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysG 88
1554 COTGCAGATGAACAGCCCTGCGTGTGAGGACACTGCCGCTCTATTATTG 1603
88 lYAlaGlyMetAsnThrAspArgGlyLys..ThrGlnLysValThrLeuAs 104
1604 CT.....CGAGGCAGCCACTATTTCGGTCACTGGCAC 1635
104 nValHisSerGluTyrGluProSerTyrGluGluGlnProMetProGluT 121
1636 TTCGCGGTGTGGGGTCAAGGAACCCCTGG.....TCACCGTCTCCTCGG 1679
121 hrProLysTrpPheHis.....Leu 127
1680 CTCACCAAGGCCCATCGGTCTCCCTCCTAGCACCCCTCTCCCAAGACA 1729
128 ProTyrLeuPheGlnMetProAlaTyrAlaSerSerSerLysPheValTh 144
1730 CCTCTGGGGGCACAGCGCCCTGGGTGCTGCTGTCA..... 1765
144 rArgValThrThrProAlaGln.....ArgGlyLysValProProv 158
1766 .AGGACTACTTCCCGAACCCGCTGACGGTGTCTGTGAACTCAGCGCCCT 1814
158 alHisHisSerSerProThrThrGlnIleThrHisArgProArgValSer 174
1815 GACCACGGCGTGCACACCTTCCCGGTGCTCTACAGTCCCTCAGGACTCT 1864
175 ArgAlaSerSerValAlaGlyAspLysProArgThrPheLeuProSerTh 191
1865 ACTCCCTCAGCAGCGGTGGTACCGTGCCT.....CCAGCAGC 1902
191 rThrAlaSerLysIleSerAlaLeuGluGlyLeuLeuLysProGlnThrP 208
1903 TTGGGCACCCAGACCTACATCTGCAACGTGAATCACAGCCCAACACAC 1952
208 roSerTyr 210
1953 CAAGGTGG 1960
```

```
seq_name: /cgn2_6/ptodata/1/lna/5B_COMB.seq:US-08-157-101A-9
seq_documentation_block:
; Sequence 9, Application US/08157101A
; Patent No. 5808032
; GENERAL INFORMATION:
; APPLICANT: KURIHARA, TATSUYA
; APPLICANT: MATSUKURA, SHIGEKAZU
; APPLICANT: TSURUOKA, NOBUO
; APPLICANT: ARIMA, KENJI
; APPLICANT: NISHIHARA, TATSURO
; TITLE OF INVENTION: ANTI-HBS ANTIBODY GENES AND EXPRESSION
; TITLE OF INVENTION: PLASMIDS THEREFOR
; NUMBER OF SEQUENCES: 9
```





Percent Similarity: 46.816 Percent Identity: 23.221

alignment\_block:

US-09-135-238B-2\_COPY\_18\_253 x US-08-463-587A-24 ..

Align seg 1/1 to: US-08-463-587A-24 from: 1 to: 2178

```
13 GlyGlySerValThrIleLysCysProLeuProGluMetHisValArg.. 28
|||||.....:.....:.....:.....:.....:.....:
907 GGGGGCTCACTCCGTTGTCTGTGACGTTCTGCTTCAACATTAAAGA 956
29 .....IleTyrLeuCysArgGluMetAlaClySerGlyThr.....C 41
|||||.....:.....:.....:.....:.....:.....:
957 CACCTATATACACTGGGTGGCTCAGGCCCGCGGTAAAGGCGCTGGAATGGG 1006
41 ySgLyThrValValSerThrThrAsnPheIleLys.....AlaGluTyr 55
|||||.....:.....:.....:.....:.....:.....:
1007 TTGCAAGATTTCCTACGAATGGTTATCTAGATATGCGGATAGCGTC 1056
56 LysGlyArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuVa 72
|||||.....:.....:.....:.....:.....:.....:
1057 AAGGCGCGTTTCACTATAAGCGAGACACATCCAAACACACAGCCCTACCT 1106
72 lGluValThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyA 89
|||||.....:.....:.....:.....:.....:.....:
1107 GCAGATGAACACAGCTGGTGTGAGGACACTGCCGTCTATTATTGTTCTA 1156
89 la.GlyMetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnVa 105
|||||.....:.....:.....:.....:.....:.....:
1157 GATGGGAGGGGAGS.....GCTTCATGCTATGGACTACTGGGGGT 1197
105 lHisSerGluTyrGluProSerTyrPgluGlnProMetProGluThrp 122
|.....:.....:.....:.....:.....:.....:
1198 CAAGAAACCTGGTCCACCGCTCTCCGCCCTCCACCAAGGCCCATCGGT 1247
122 roLysTyrPheHisLeuProTyrLeuPheGlnMetProAlaTyrAlaser 138
|||.....:.....:.....:.....:.....:.....:
1248 CTTCCCCCTGGCACCTCTCCAGAGCACCTCTGGGGGCACAGCGGCC 1297
139 SerSerLysPheValThrArgValThrProAlaGln.....Ar 152
|||.....:.....:.....:.....:.....:.....:
1298 TGGGTGCTCGG...TCAAGGACTACTTCCCGCAACCGGTGACGTGTGC 1344
152 gGlyLysValProValHisHisSerSerProThrThrGlnIleThrH 169
|||||.....:.....:.....:.....:.....:.....:
1345 TGGAACTCAGCGCCCTGACCAGCGCGGTGCACACCTCCCGGTGTCT 1394
169 isArgProArgValSerArgAlaSerSer.....Val 179
|||.....:.....:.....:.....:.....:.....:
1395 ACAGTCCTCAGGACTCTACTCCCTCAGCAGCGTGGTACTGTGCCCTCTA 1444
180 AlaGlyAspLysProGluThrPheLeuProSerThrThrAlaSerLysII 196
|||||.....:.....:.....:.....:.....:.....:
1445 GCAGCTTGGGACCCAGA.....CCTACATCTGCAACGTGAAT 1482
196 eSerAlaLeuGluGlyLeuLysProGlnThrPro..... 208
|||.....:.....:.....:.....:.....:.....:
1483 CACAAGC.....CCAGCACACCAAGGTGGACAGAA 1514
209 .....SerTyrAsnHisThrArgLeu.HisArgGlnArgAlaLeu.A 223 *
|||.....:.....:.....:.....:.....:.....:
1515 AGTTAGCCCAATCTTGTGCACAAACTCACAGGGCCCTTCGTTTGTG 1564
223 spTyr..... 224
|||.....:.....:.....:.....:.....:.....:
1565 AATATCAGGCCAATCGTCTGACCTGCCTCAACCTCTCTCAATGCTGGC 1614
225 .....GlySerGlnSerGlyArgGluGlyGlnGly 234
|||||.....:.....:.....:.....:.....:.....:
1615 GCGGCTCTGGTGGTGTCTGGTGGCGGCTCTGAGGGTGGTGGC 1659
```



OM of: US-09-135-238b-2\_COPY\_18\_272 to: GenEmbl:\* out\_format : pfs

Date: Sep 12, 2000 6:55 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODE=frame+p2n.model -DEV=xlp  
-Q/cgn2\_1/USPIO.spool/US09135238/runat\_24082000\_091323\_17880/app\_query.fasta\_1.1282  
-DB=GenEmbl -OFMT=fastap -SURFIX=p2n.rge -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-CGAPOP=4.500 -CGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=ppct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM\_ext -MINLEN=0 -MAXLEN=1000000  
-USER=US09135238 @CGN1\_1\_5120 -NCPU=6 -ICPU=3 -LONGLOG -NO\_XLPXY  
-WAIT -THREADS=1

Search information block:

Query: US-09-135-238b-2\_COPY\_18\_272

Query length: 255

Database: GenEmbl.\*

Database sequences: 972840

Database length: 892348106

Search time (sec): 2784.300000

score\_list:

Sequence	Strtd	Orig	EScore	Len	Documentation
gb_pr3:AF057557	+	1340.00	1822.33	2.1e-93	1339
gb_htg11:AC023534	+	629.50	807.63	6.9e-37	221365
gb_htg22:AC027719	-	252.00	304.25	7.5e-09	58628
gb_to:MMU06431	+	197.00	255.20	4.1e-06	3095
gb_cm:AF091137	+	189.00	245.02	1.5e-05	2848
gb_to:MMP1GR03	+	188.00	260.64	2.0e-06	417
gb_to:MMP1GR03	+	188.00	258.94	2.5e-06	505
gb_to:RNRG	+	188.00	232.02	0.0003	29392
gb_to:RNRG	+	185.50	239.02	0.0003	29392
gb_to:RNRG	+	181.00	235.39	3.1e-05	3269
gb_cm:AB032195	+	181.00	235.39	3.1e-05	2461
gb_pr2:S43433503	+	180.50	232.08	6.1e-06	345
gb_pr2:S43433503	+	180.50	236.83	4.3e-05	1936
gb_pr2:S43433503	+	180.50	236.83	4.3e-05	2031
gb_pr2:S43433503	+	180.50	234.91	5.5e-05	2405
gb_pr2:S43433503	+	180.50	233.26	6.8e-05	2898
gb_pr1:HSP1R	+	180.50	233.20	6.8e-05	2919
gb_cm:BP1GR1	+	179.50	239.91	0.0001	3630
gb_cm:BP1GR1	+	179.00	229.48	0.0001	3527
gb_cm:BP1GR1	+	175.50	230.12	0.0001	1911
gb_cm:OC1GR1	+	166.50	212.45	0.0010	3517
gb_cm:OC1GR1	+	166.00	175.14	0.1168	221365
gb_htg11:AC023534	-	146.50	203.22	0.0032	456
gb_to:MMP1GR5	+	143.00	193.31	0.0410	2526
gb_pr1:HUM1GRPOLY	+	137.00	189.78	0.0179	481
gb_to:MMP1GR5	+	132.50	186.89	0.0259	333
gb_pr2:S43433505	+	130.00	177.47	0.0867	657
gb_pr2:S43433504	+	124.00	165.85	0.3848	969
gb_pr2:HS4238323	+	124.00	165.84	0.3853	970
gb_pr2:HS4238323	+	124.00	165.10	0.8046	1857
gb_pr2:HS4238323	+	121.00	135.35	69.33	59506
gb_htg29:AC060797	-	121.00	135.35	69.33	175260
gb_htg15:AC016888	-	118.50	149.83	3.00	2539
gb_to:AF035963	+	118.50	149.83	3.00	1440
gb_pr3:AF0433724	+	117.50	153.48	1.88	1351
gb_cm:RAB1CBXB	+	114.50	149.95	2.96	16288
gb_pr1:HSNCAMX1	+	111.00	132.15	91.83	2799
gb_pr1:HSNCAMX1	+	110.50	138.06	13.61	251700
gb_ba1:AP000062	+	110.50	98.27	2.2e+03	58628
gb_htg22:AC027719	+	109.50	109.79	510.16	175799
gb_htg5:AC013305	-	109.50	100.08	1.8e+03	175799

gb\_pr2:HUMTCBYZ + 109.00 144.47 5.97 1074 ! M15564 Human T-cell recepto  
gb\_pr2:HUMTCBYZ + 108.50 145.00 5.58 936 ! L34734 Homo sapiens T-cell r  
gb\_htg10:AC024170 - 108.00 97.86 2.4e+03 179293 ! AC024170 Homo sapiens chr  
gb\_htg28:AC021643 - 108.00 94.81 3.5e+03 253327 ! AC021643 Mus musculus chr  
gb\_ro:MMP1GR04 + 107.50 145.70 5.10 741 ! U83429 Mus musculus polymeri  
seq\_name: gb\_pr3:AF057557

seq\_documentation\_block: 1339 bp mRNA PRI 04-JUN-1998  
LOCUS AF057557  
DEFINITION Homo sapiens anti-Fas-induced apoptosis (TOSO) mRNA, complete cds.  
ACCESSION AF057557  
VERSION AF057557.1 GI:3169292  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1339)

AUTHORS Hitoshi,Y., Lorens,J., Kitada,S.I., Fisher,J., LaBarge,M.,  
Ring,H.Z., Francke,U., Reed,J.C., Kinoshita,S. and Nolan,G.P.  
TITLE Toso, a cell surface, specific regulator of Fas-induced apoptosis  
in T cells

JOURNAL Immunity 8 (4), 461-471 (1998)  
MEDLINE 98246046  
REFERENCE 2 (bases 1 to 1339)  
AUTHORS LaBarge,M. and Hitoshi,Y.  
TITLE Direct Submission  
JOURNAL Submitted (03-APR-1998) Molecular Pharmacology, Stanford  
University, 300 Pasteur Dr., Palo Alto, CA 94304, USA

FEATURES  
Location/Qualifiers  
1..1339  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="ig31-q32"  
/cell\_lines="Jurkat"  
/cell\_type="T cell"  
/note="TOSO, 2 CDNA: cloned by PCR using primers designed  
from the original TOSO.1 cDNA sequence"  
1..1339  
/gene="TOSO"  
20..1192  
/gene="TOSO"  
/note="transmembrane protein; Ig superfamily member"  
/codon\_start=1  
/product="anti-Fas-induced apoptosis"  
/protein\_id="AAC18830.1"  
/db\_xref="GI:3169293"

gene  
CDS  
variation

variation  
BASE COUNT 309 a 432 c 340 g 258 t  
ORIGIN

alignment\_scores:  
Quality: 1340.00 Length: 255  
Ratio: 5.255 Gaps: 0  
Percent similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-135-238b-2\_copy\_18\_272 x AF057557 ..

Align seg 1/1 to: AF057557 from: 1 to: 1339

```

1 ArgIleLeuProGluValLysValGluGluLeuGlyGlySerValTh 17
|||||
71 AGGATCCTCCAGAGTAAAGGTAGAGGGGAGCTGGCGGATCATTC 120
|||||
17 rIleLysCysProLeuProGluMetHisValArgIleTyLeuCysArgG 34
|||||
121 CATCAAGTCCCATCTCTCGAATGCATGTGAGGATATATCTGTCCGGG 170
|||||
34 luMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsnPhe 50
|||||
171 AGATGGCTGGATCTGGAACATGTGTGTCGTGTATCCACCACTTC 220
|||||
51 IleLysAlaGluTyLysGlyArgValThrLeuLysGlnTyProArgLy 67
|||||
221 ATCAAGCCAGATACAGAGCCGAGTTACTCTGAGCATATCCACGCAA 270
|||||
67 sAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerGly 84
|||||
271 GAATCTGTCTCTAGTGGAGTAAACAGCTGACAGAAAGTGACAGCGAG 320
|||||
84 alTyAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLys 100
|||||
321 TCTATGCTCGGAGCGGCATGACACAGACCGGGGAAGACCCAGAA 370
|||||
101 ValThrLeuAsnValHisSerGluTyArgGluProSerTrpGluGluInP 117
|||||
371 GTACCCCTGAATGTCCACAGTGAATACGAGCCATCATGGGAAGACGAC 420
|||||
117 oMetProGluThrProLysTrpPheHisLeuProTyLeuPheGlnMetP 134
|||||
421 AATGCTCGAGACTCCAAATGGTTTCATCTGCCCTATTGTTCAGATGC 470
|||||
134 roAlaTyAlaSerSerSerLysPheValThrArgValThrThrProAla 150
|||||
471 CTGCATATGCCAGTCTCTCCAAATTCGTAACCAAGTTACACACAGCT 520
|||||
151 GlnArgGlyLysValProProValHisHisSerSerProThrThrGlnI 167
|||||
521 CAAGGGGCAAGTCCCTCCAGTTCACACCTCTCCCCACACCCCAAT 570
|||||
167 eThrHisArgProArgValSerArgAlaSerSerValAlaGlyAspLysP 184
|||||
571 CACCCACCGCCCTCGAGTGTCCAGAGCATCTTCAGTAGCAGGTGCAAGC 620
|||||
184 roArgThrPheLeuProSerThrThrAlaSerLysIleSerAlaLeuGlu 200
|||||
621 CCGGAACCTCTCTGCCATCCACTACAGCTCCAAATCTCAGCTCTGGAG 670
|||||
201 GlyLeuLeuLysProGlnThrProSerTyHisHisThrArgLeuHi 217
|||||
671 GGCTGCTCTCAAGCCCCAGAGCCAGCTTACAAACCCACACACAGGCTGCA 720
|||||
217 sArgGlnArgAlaLeuAspTyArgLysGlnSerGlyArgGluGlyGlnG 234
|||||
721 CAGGCAGAGACACTGGACTATGGCTACAGCTCTGGGAGGAAGGCCAAG 770
|||||
234 lyPheHisIleLeuIleProThrIleLeuGlyLeuPheLeuAlaLeu 250
|||||
771 GATTCATCATCTGATCCGAGCACCCTCTGGGCCCTTTTCCTGCTGGCACTT 820
|||||
251 LeuGlyLeuValVal 255
|||||
821 CTGGGGCTGGTGGT 835
|||||
seq_name: gb_htg11:AC023534
seq_documentation_block:
LOCUS AC023534 221365 bp DNA HTG 02-MAR-2000

```

## DEFINITION

Homo sapiens clone RP11-462N18, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 61

unordered pieces.

## ACCESSION

AC023534.2 GI:7143452

## VERSION

HTG: HTGS\_PHASE1.

## KEYWORDS

human.

## SOURCE

Homo sapiens

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 221365)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Homo sapiens chromosome, clone RP11-462N18

Unpublished

2 (bases 1 to 221365)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,

Boukhalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,

Choepe,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,

DeArelano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,

Fenster,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,

Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,

Grand-pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,

Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,

Klein,J., Landers,T., Larcocque,K., Lehoczy,J., Levine,R.,

Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,

McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrum,J.,

Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,

Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olivari,T.M.,

Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,

Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,

Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,

Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tirrell,A.,

Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,

Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and

Zody,M.

Direct Submission

Submitted (15-FEB-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Mar 2, 2000 this sequence version replaced gi:6978210.

All repeats were identified using RepeatMasker:

Smit, A.F.A. &amp; Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L524

Center clone name: 462\_N\_18

-----

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 61 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 1108: contig of 1108 bp in length

\* 1109 1208: gap of 100 bp

\* 1209 2292: contig of 1084 bp in length

\* 2293 2392: gap of 100 bp

\* 2393 3673: contig of 1281 bp in length

\* 3674 3773: gap of 100 bp

\* 3774 4947: contig of 1174 bp in length

\* 4948 5047: gap of 100 bp

\* 5048 6104: contig of 1057 bp in length

\* 6105 6204: gap of 100 bp

\* 6205 7259: contig of 1055 bp in length

\* 7260 7359: gap of 100 bp

\* 7360 8667: contig of 1308 bp in length

\* 8668 8767: gap of 100 bp

```
*      8768      10566: contig of 1799 bp in length
*      10567      10666: gap of 100 bp
*      10867      12173: contig of 1507 bp in length
*      12174      12273: gap of 100 bp
*      12274      13643: contig of 1370 bp in length
*      13644      13743: gap of 100 bp
*      13744      15215: contig of 1472 bp in length
*      15216      15315: gap of 100 bp
*      15316      16609: contig of 1294 bp in length
*      16610      16709: gap of 100 bp
*      16710      18085: contig of 1376 bp in length
*      18086      18185: gap of 100 bp
*      18186      19668: contig of 1483 bp in length
*      19669      19768: gap of 100 bp
*      19769      22083: contig of 2315 bp in length
*      22084      22183: gap of 100 bp
*      22184      24052: contig of 1869 bp in length
*      24053      24152: gap of 100 bp
*      24153      25902: contig of 1750 bp in length
*      25903      26002: gap of 100 bp
*      26003      27639: contig of 1637 bp in length
*      27640      27739: gap of 100 bp
*      27740      29377: contig of 1638 bp in length
*      29378      29477: gap of 100 bp
*      29478      30910: contig of 1433 bp in length
*      30911      31010: gap of 100 bp
*      31011      31225: contig of 215 bp in length
*      31226      31325: gap of 100 bp
*      31326      32890: contig of 1565 bp in length
*      32891      32990: gap of 100 bp
*      32991      34812: contig of 1822 bp in length
*      34813      34912: gap of 100 bp
*      34913      36817: contig of 1905 bp in length
*      36818      36917: gap of 100 bp
*      36918      38575: contig of 1658 bp in length
*      38576      38675: gap of 100 bp
*      38676      41309: contig of 2634 bp in length
*      41310      41409: gap of 100 bp
*      41410      43344: contig of 1935 bp in length
*      43345      43444: gap of 100 bp
*      43445      45853: contig of 2409 bp in length
*      45854      45953: gap of 100 bp
*      45954      48434: contig of 2481 bp in length
*      48435      48534: gap of 100 bp
*      48535      51026: contig of 2492 bp in length
*      51027      51126: gap of 100 bp
*      51127      54369: contig of 3243 bp in length
*      54370      54469: gap of 100 bp
*      54470      56661: contig of 2192 bp in length
*      56662      56761: gap of 100 bp
*      56762      59097: contig of 2336 bp in length
*      59098      59197: gap of 100 bp
*      59198      61551: contig of 2354 bp in length
*      61552      61651: gap of 100 bp
*      61652      64257: contig of 2606 bp in length
*      64258      64357: gap of 100 bp
*      64358      67511: contig of 3154 bp in length
*      67512      67611: gap of 100 bp
*      67612      70329: contig of 2718 bp in length
*      70330      70429: gap of 100 bp
*      70430      73572: contig of 3143 bp in length
*      73573      73672: gap of 100 bp
*      73673      77203: contig of 3531 bp in length
*      77204      77303: gap of 100 bp
*      77304      80599: contig of 3296 bp in length
*      80600      80699: gap of 100 bp
*      80700      84046: contig of 3347 bp in length
*      84047      84146: gap of 100 bp
*      84147      88357: contig of 4211 bp in length
*      88358      88457: gap of 100 bp
*      88458      92454: contig of 3997 bp in length
*      92455      92554: gap of 100 bp
*      92555      96892: contig of 4338 bp in length
```

```
*      96893      96992: gap of 100 bp
*      96993      101334: contig of 4342 bp in length
*      101335      101434: gap of 100 bp
*      101435      105438: contig of 4004 bp in length
*      105439      105538: gap of 100 bp
*      105539      110413: contig of 4875 bp in length
*      110414      110513: gap of 100 bp
*      110514      115994: contig of 5481 bp in length
*      115995      116094: gap of 100 bp
*      116095      120447: contig of 4353 bp in length
*      120448      120547: gap of 100 bp
*      120548      124483: contig of 3936 bp in length
*      124484      124583: gap of 100 bp
*      124584      130249: contig of 5666 bp in length
*      130250      130349: gap of 100 bp
*      130350      135462: contig of 5113 bp in length
*      135463      135562: gap of 100 bp
*      135563      139847: contig of 4285 bp in length
*      139848      139947: gap of 100 bp
*      139948      146731: contig of 6784 bp in length
*      146732      146831: gap of 100 bp
*      146832      151469: contig of 4638 bp in length
*      151470      151569: gap of 100 bp
*      151570      159901: contig of 8332 bp in length
*      159902      160001: gap of 100 bp
*      160002      168308: contig of 8307 bp in length
*      168309      168408: gap of 100 bp
*      168409      176195: contig of 7787 bp in length
*      176196      176295: gap of 100 bp
*      176296      189303: contig of 13008 bp in length
*      189304      189403: gap of 100 bp
*      189404      202135: contig of 12732 bp in length
*      202136      202235: gap of 100 bp
*      202236      221365: contig of 19130 bp in length.

FEATURES             Location/Qualifiers
     source            1..221365
                       /organism="Homo sapiens"
                       /db_xref="taxon:9606"
                       /clone_lib="RPC1-II Human Male BAC"
     misc_feature       1..1108
                       /note="assembly_fragment"
     misc_feature      1209..2292
                       /note="assembly_fragment"
     misc_feature      2393..3673
                       /note="assembly_fragment"
     misc_feature      3774..4947
                       /note="assembly_fragment"
     misc_feature      5048..6104
                       /note="assembly_fragment"
     misc_feature      6205..7259
                       /note="assembly_fragment"
     misc_feature      7360..8667
                       /note="assembly_fragment"

alignment_scores:
    Quality: 629.50      Length: 476
    Ratio: 3.459        Gaps: 6
    Percent Similarity: 38.235      Percent Identity: 34.244

alignment_block:
US-09-135-238B-2_COPY_18_272 x AC023534      ..

Align seg 1/1 to: AC023534 from: 1 to: 221365

1 ArgileLeuProGluValLysValGluGlyLeuLeuGlySerValTh 17
|||||
33190 AGGATCTCTCCAGAGTAAAGGTAGAGGGGAGCTGGCGGATCAGTTAC 33239

17 rIleLysCysProLeuProGluMetHisValArgIleTyrLeuCysArg 34
|||||
33240 CATCAAGTGCCCACTTCCTGAAATGCATGTGAGGATATATCTGTGCGGG 33289

34 luMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsnPhe 50
```

```
|||||
33290 AGATGGCTGGATCTGGAAATGTGTACCGTGGTATCCACCAACTTC 33339
51 IleLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProArgLy 67
|||||
33340 ATCAGGCAGATATACAGGCGCGAGTTACTCTGAGCAATATCCACGCAA 33389
67 sAsnLeuPheLeuValGlnValThrGlnLeuThrGluSerAspSerGlyV 84
|||||
33390 GAATCTGTCTCTAGTGGAGGTAAACACAGCTGACAGAAAGTGACAGCGGAG 33439
84 alTyAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLys 100
|||||
33440 TCTATGCTTCGGAGCGGCATGAACACAGACCGGGGAAGACCCAGAAA 33489
101 ValThrLeuAsnValHis..... 106
|||||
33490 GTCACTGTAATGTGCACAGCTGGTAGTTCTCCGCTGATTGGAGGCTCA 33539
106 ..... 106
33540 GACTACCCAGAAATGTTTCACTTGGGAAGATAGGAACAGACATCTACT 33589
106 ..... 106
33590 GTGACAGAGTCTAGACTCTTACACTGTGCCATATAAGCCACATAATGA 33639
106 ..... 106
33640 TTGAGATCCATGAACCTCATAGTCAATTACTCTCATGTCGTCACACAA 33689
106 ..... 106
33690 TTTTCTACCAACATGTATGAAAGAACAGATAGGTGCTCTCTTTGTAT 33739
106 ..... 106
33740 TGTTCATCTACCCAGGCATGAGGGAGTAGAGGTTGAGATGGAGAG 33789
106 ..... 106
33790 AGGCTCTGAGATGGAGGAATTGGAAGCAAAAGAAATAACTATCCAAAG 33839
106 ..... 106
33840 TAATTGTCCTATGAGAAATAAGGAAGACACTAATTATCCAGCCACAT 33889
106 ..... 106
33890 TTATAGCCAAATCTTATTACTCTGTGACCTGATCCCTTTATTCTCGAA 33939
106 ..... 106
33940 CCCCCTTTCTCTTCTTCTCCACTGAGTCCCAACACATATGACAGAGT 33989
106 ..... 106
33990 AATTCAACAGCCTAGATCCCAATGACCTCTGCTGTTTTTTCAGGGAATG 34039
106 ..... 106
34040 TTGACTGGACTAGGGNACCAGTATTTCTCCAGGGGAAGCCCAACCCA 34089
106 ..... 106
34090 GCTCCTTCTCTCTTCTTCTTCTAGTGACCTTGCCCTCATCCCTCACTGTC 34139
106 ..... 106
34140 ATCCACTTTGCTGAATAGCTTCAAGTTTACACCAAGCCTCAGTCFCCC 34189
107 .....Ser..GluTyrGluProS 112 *
|||||
```

```
34190 CCTAGAGACCACCAACATATTCTCTCTTCCATAGAATAAGACGCCAT 34239
112 erTTPGluGlnProMetProGluThrProLysTyrPheHisLeuPro 128
|||||
34240 CATGGGAAGACAGCCATGCCAGACTCCAAATGGTTTCATCTGCCC 34289
129 TyrLeuPheGlnMetProAlaTyrAlaSerSerSerLysPheValThrAr 145
|||||
34290 TATTGTTCAGATGCTTCATATGCGAGTTCTTCCAAATTCGTAACCCAG 34339
145 g...ValThrThrProAlaGlnArgGlyLysValPro..... 156
|||||
34340 AGGTCAGTTCACCATAGGAGGGTAGGAGGTGGGTGAGATGCTGTGCC 34389
157 .....ProValHisHisSerSerPro 163
|||||
34390 AATGTGAAATGGAGGGGTCTATGTTTCAGTTTCAT..... 34427
164 ThrThrGlnIleThrHisArgProArgValSerArgAlaSerSerValAl 180
|||||
34428 .....AGATGGCTGGCAGAGCTGAGCTCTAGGTGTGTATTCCTC 34471
180 aGlyAspLysProArgThrPheLeuProSerThrThrAlaSerLysIleS 197
|||||
34472 TGCCTTTAGAGAGGGCACATACCTCCCTGAGATCTGCGGTCCATATCCTT 34521
197 erAlaLeuGluGlyLeuLeuLysProGlnThrProSerTyrAsnHisHis 213
|||||
34522 CAAGCTTCCACAGACTCAGTCCCTGCTCTAGCTGCAGAACTTACAT 34571
214 .....ThrArgLeuHisArgGlnArg 220
|||||
34572 ATTACCACGAGGTGGCACCAACAGCAA 34598
seq_name: gb_hgt22.AC027719
seq_documentation_block:
LOCUS AC027719 58628 bp DNA HTG 01-APR-2000
DEFINITION Homo sapiens chromosome 8 clone RP11-258P9 map 8, LOW-PASS SEQUENCE
SAMPLING.
ACCESSION AC027719
VERSION AC027719.1 GI:7382531
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 8, clone RP11-258P9
JOURNAL Unpublished
REFERENCE
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Headford,A., Horton,L.,
Hewland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mienda,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
```

TITLE  
JOURNAL  
COMMENT

Direct Submission  
Submitted (01-APR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L7449  
Center clone name: 258\_P\_9  
-----

\* NOTE: This record contains 74 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 671: contig of 671 bp in length  
672 771: gap of 100 bp  
772 1469: contig of 698 bp in length  
1470 1569: gap of 100 bp  
1570 2261: contig of 692 bp in length  
2262 2361: gap of 100 bp  
2362 3051: contig of 690 bp in length  
3052 3151: gap of 100 bp  
3152 3816: contig of 665 bp in length  
3817 3916: gap of 100 bp  
3917 4597: contig of 681 bp in length  
4598 4697: gap of 100 bp  
4698 5395: contig of 698 bp in length  
5396 5495: gap of 100 bp  
5496 6185: contig of 690 bp in length  
6186 6285: gap of 100 bp  
6286 6969: contig of 684 bp in length  
6970 7069: gap of 100 bp  
7070 7780: contig of 711 bp in length  
7781 8590: contig of 710 bp in length  
8591 8690: gap of 100 bp  
8691 9372: contig of 682 bp in length  
9373 9472: gap of 100 bp  
9473 10162: contig of 690 bp in length  
10163 10262: gap of 100 bp  
10263 10944: contig of 682 bp in length  
10945 11044: gap of 100 bp  
11045 11747: contig of 703 bp in length  
11748 11847: gap of 100 bp  
11848 12547: contig of 700 bp in length  
12548 12647: gap of 100 bp  
12648 13336: contig of 689 bp in length  
13337 13436: gap of 100 bp  
13437 14103: contig of 667 bp in length  
14104 14203: gap of 100 bp  
14204 14872: contig of 669 bp in length  
14873 14972: gap of 100 bp  
14973 15682: contig of 710 bp in length  
15683 15782: gap of 100 bp  
15783 16491: contig of 709 bp in length  
16492 16591: gap of 100 bp  
16592 17290: contig of 699 bp in length  
17291 17390: gap of 100 bp  
17391 18088: contig of 698 bp in length  
18089 18188: gap of 100 bp  
18189 18883: contig of 695 bp in length

\* 18884 18983: gap of 100 bp  
18884 19671: contig of 688 bp in length  
19672 19771: gap of 100 bp  
19772 20413: contig of 642 bp in length  
20414 20513: gap of 100 bp  
20514 21196: contig of 683 bp in length  
21197 21296: gap of 100 bp  
21297 21992: contig of 696 bp in length  
21993 22092: gap of 100 bp  
22093 22804: contig of 712 bp in length  
22805 22904: gap of 100 bp  
22905 23604: contig of 700 bp in length  
23605 23704: gap of 100 bp  
23705 24402: contig of 698 bp in length  
24403 24502: gap of 100 bp  
24503 25191: contig of 689 bp in length  
25192 25291: gap of 100 bp  
25292 25974: contig of 683 bp in length  
25975 26074: gap of 100 bp  
26075 26769: contig of 695 bp in length  
26770 26869: gap of 100 bp  
26870 27572: contig of 703 bp in length  
27573 27672: gap of 100 bp  
27673 28373: contig of 701 bp in length  
28374 28473: gap of 100 bp  
28474 29165: contig of 692 bp in length  
29166 29265: gap of 100 bp  
29266 29956: contig of 691 bp in length  
29957 30056: gap of 100 bp  
30057 30741: contig of 685 bp in length  
30742 30841: gap of 100 bp  
30842 31519: contig of 678 bp in length  
31520 31619: gap of 100 bp  
31620 32323: contig of 704 bp in length  
32324 32423: gap of 100 bp  
32424 33110: contig of 687 bp in length  
33111 33210: gap of 100 bp  
33211 33902: contig of 692 bp in length  
33903 34002: gap of 100 bp  
34003 34708: contig of 706 bp in length  
34709 34808: gap of 100 bp  
34809 35500: contig of 692 bp in length  
35501 35600: gap of 100 bp  
35601 36300: contig of 700 bp in length  
36301 36400: gap of 100 bp  
36401 37097: contig of 697 bp in length  
37098 37197: gap of 100 bp  
37198 37900: contig of 703 bp in length  
37901 38000: gap of 100 bp  
38001 38695: contig of 695 bp in length  
38696 38795: gap of 100 bp  
38796 39487: contig of 692 bp in length  
39488 39587: gap of 100 bp  
39588 40291: contig of 704 bp in length  
40292 40391: gap of 100 bp  
40392 41101: contig of 710 bp in length  
41102 41201: gap of 100 bp  
41202 41886: contig of 685 bp in length  
41887 41986: gap of 100 bp  
41987 42686: contig of 700 bp in length  
42687 42786: gap of 100 bp  
42787 43484: contig of 698 bp in length  
43485 43584: gap of 100 bp  
43585 44276: contig of 692 bp in length  
44277 44376: gap of 100 bp  
44377 45078: contig of 702 bp in length  
45079 45178: gap of 100 bp  
45179 45869: contig of 691 bp in length  
45870 45969: gap of 100 bp  
45970 46664: contig of 695 bp in length  
46665 46764: gap of 100 bp  
46765 47456: contig of 692 bp in length  
47457 47556: gap of 100 bp

```

* 47557 48249: contig of 693 bp in length
* 48250 48349: gap of 100 bp
* 48350 49042: contig of 693 bp in length
* 49043 49142: gap of 100 bp
* 49143 49843: contig of 701 bp in length
* 49844 49943: gap of 100 bp
* 49944 50640: contig of 697 bp in length
* 50641 50740: gap of 100 bp
* 50741 51438: contig of 698 bp in length
* 51439 51538: gap of 100 bp
* 51539 52231: contig of 693 bp in length
* 52232 52331: gap of 100 bp
* 52332 53020: contig of 689 bp in length
* 53021 53120: gap of 100 bp
* 53121 53857: contig of 737 bp in length
* 53858 53957: gap of 100 bp
* 53958 54652: contig of 695 bp in length
* 54653 54752: gap of 100 bp

```

```

alignment_scores:
  Quality: 252.00      Length: 48
  Ratio: 5.250        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

```

## alignment\_block:

US-09-135-238B-2\_COPY\_18\_272 x AC027719/rev ..

Align seg 1/1 to reverse of: AC027719 from: 1 to: 58628

```

1 Argf1eLeuProGluValLysValGluGlyLeuGlySerValTh 17
|||||
40580 AGGATCTCCCAAGTAAGAGTAGAGGGGAGCTGGCGGATCATTTAC 40531

```

```

17 r1eLysCysProLeuProGluMetHisValArgIleTyrlenCysArg 34
|||||
40530 CATCAAGTGCACATCTCTGGAATGCATGTGAGGATATCTGTGCGGG 40481

```

```

34 luMetAlaGlySerGlyThrCysGlyThrValValSerThrThr 48
|||||
40480 AGATGGCTGGATCTGGAACATGTGTACCGTGTATCCACCACC 40437

```

seq\_name: gb\_ro:MMU06431

```

seq_documentation_block:
LOCUS      MMU06431      3095 bp      mRNA      ROD      26-MAY-1995
DEFINITION Mus musculus B6/CBAF1J polymeric immunoglobulin receptor mRNA,
complete cds.

```

```

ACCESSION      U06431
VERSION        U06431.1
KEYWORDS       GI:458421
SOURCE         mouse.

```

```

ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```

REFERENCE      Piskurich,J.F., Blanchard,M.H., Youngman,K.R., France,J.A. and
AUTHORS        Kaetzel,C.S.
TITLE          Molecular cloning of the mouse polymeric Ig receptor. Functional
regions of the molecule are conserved among five mammalian species
J. Immunol. 154 (4), 1735-1747 (1995)

```

```

JOURNAL        J. Immunol. 154 (4), 1735-1747 (1995)
MEDLINE        95138517
REFERENCE      Piskurich,J.F., Hsieh,M., Youngman,K.R. and Kaetzel,C.S.
AUTHORS        Piskurich,J.F., Hsieh,M., Youngman,K.R. and Kaetzel,C.S.
TITLE          Molecular cloning of mouse polymeric Ig receptor
JOURNAL        J. Immunol. 150, 38A-38A (1993)
REFERENCE      3 (bases 1 to 3095)
AUTHORS        Piskurich,J.F.
TITLE          Molecular Cloning and Regulation of the Polymeric Immunoglobulin
Receptor

```

```

JOURNAL        Thesis (1994) Pathology, Case Western Reserve University
REFERENCE      4 (bases 1 to 3095)
AUTHORS        Piskurich,J.F.
TITLE          Direct Submission

```

JOURNAL Submitted (08-FEB-1994) Janet F. Piskurich, Case Western Reserve University, Pathology, Biomedical Research Building, Cleveland, OH 44060, USA

## FEATURES

```

source
  1..3095
    Location/Qualifiers
      /organism="Mus musculus"
      /strain="B6/CBAF1J (C57 Black 6 X CBA)"
      /db_xref="taxon:10090"
      /clone="G20"
      /clone_lib="catalog #935302, Stratagene Cloning Systems,
      La Jolla, CA"
      /sex="female"
      /tissue_type="liver"
      <1..84
        5'UTR
      85..138
        sig_peptide
      85..2400
        CDS
      /codon_start=1
      /function="binds and transports polymeric immunoglobulin"
      /product="polymeric immunoglobulin receptor"
      /protein_id="AAA67440.1"
      /db_xref="GI:458422"
      /translation="MRLYLFTLLVTFVSGVSTKSPIFGPQEVSSIIEGDSVITCVYPD
      TSVNRHTRKYKCGAGSMCTLISSNGYLSKEYSGRANLINFPENNTFVINEOLTQ
      DDTSYKCGLGTSNGLSFDVSLVSQVPELSDHYTKDIGRNTVILCPKRENVP
      SKSLCKRTNQSCELVIDSTEKVPSTIGRAKLFMKGIDLLTFYVNIHSHLHNDAGLY
      TCDVINTLGRDPDFEGRILITPKDDNGRFSYLITGLRKEDAGHYOCGASHSGLPQE
      GWP1QTQLFVNEESTIPNRRSVKVGTVGSVALIACPYNPKSSSLKYCWKEGGDNG
      HCPALVGTQAOVEYEGRLALFDQPGNGTYVILNQLTSDAGFYWCLTNGDSRWRT
      TLEQVAETREPNTVTPONATVLTGETFTYSCHPKCYFSOKYKWCWSNKGCHIL
      PSHDEGARQSSVSCDSSQLMTPNPKSDEGHWGCVKGQOTYGETTAYIAVEE
      RTRGSSHNVTIDANAKVALEEVVDSSISEKENKAIPNPGFANERETQNVROQAO
      ENRAGGAGSADGQSRSSSKVLFSTLVLPGVLVAGAIIVAVRYRHRKNDVRMSIS
      SYRTDISMADFKNSRDGNDNMGSPTDQTVIEGKDEIVITTTCTAEPEESKAKAR
      SSKEEADWYSAFLQSSITAAQVHDGPQA"
      mat_peptide
      139..2397
        misc_feature
      /note="ectoplasmic region, binds polymeric immunoglobulin"
      2020..2088
        misc_feature
      /note="transmembrane segment of the receptor"
      2089..2397
        misc_feature
      /note="cytoplasmic tail of the receptor"
      2401..3095
        3'UTR
      BASE COUNT      861 a 796 c 784 g 654 t
      ORIGIN

```

## alignment\_scores:

```

  Quality: 197.00      Length: 148
  Ratio: 2.165        Gaps: 5
  Percent Similarity: 61.486  Percent Identity: 32.432

```

## alignment\_block:

US-09-135-238B-2\_COPY\_18\_272 x MMU06431 ..

Align seg 1/1 to: MMU06431 from: 1 to: 3095

```

13 GlyGlySerValThrIleLysCysProLeuProGlu.....Me 25
|||||
181 GCGGACTCTGTCTCCATCAGTGTACTACCCAGACACCTCTGTCAACCG 230
25 tHisValArgIleTyrlenCysArgGluMetAlaGlySerGlyThrCysG 42
|||||
231 GCACACCGGAAATACTGGTGGCGCAAGAGGCC...AGCGGATGTGCA 277
42 lyThrValValSerThrThrAsnPhelIleLysAlaGluTyrlsGlyArg 58
|||||
278 CAACGCTCATCTCTCAATGGTACCTCTCAAGAGAGTATTCAGGCAGA 327
59 ValThrLeuLysGlnTyrrProArgLysAsnLeuPheLeuValGluValTh 75
|||||
328 GCCAACCTCATCACTCCCGAGAGAACACACATTTGTGATTAACATGA 377

```



TITLE Genomic cloning and structural analysis of the murine polymorphic  
receptor (pigr) gene and promoter region  
JOURNAL Gene 201 (1-2), 189-197 (1997)  
MEDLINE 98072444  
REFERENCE 2 (bases 1 to 417)  
AUTHORS Martin.M.G., Gutierrez,E.M., Lam.J.T., Li.T.W.H. and Wang,J.  
TITLE Direct Submission  
JOURNAL Submitted (31-DEC-1996) Pediatrics, UCLA School of Medicine, 10833  
Le Conte Ave. Los Angeles, CA 90095, USA  
FEATURES Location/Qualifiers  
source  
1. .417  
/organism="Mus musculus"  
/strain="129SVJ"  
/db\_xref="taxon:10090"  
43. .387  
/gene="pigr"  
/number=3  
BASE COUNT 105 a 120 c 102 g 90 t  
ORIGIN

alignment\_scores:  
Quality: 188.00 Length: 97  
Ratio: 2.648 Gaps: 3  
Percent Similarity: 73.196 Percent Identity: 40.206

## alignment\_block:

US-09-135-238B-2\_COPY\_18\_272 x MMPIGR03 ..

Align seg 1/1 to: MMPIGR03 from: 1 to: 417

```

13 GlyGlySerValThrLeuLysCysProLeuProGlu.....Me 25
|||||
96 GGGAGCTCTGTTCCATCAGCTGCTACTACCCAGACACCTCTGTCAACCG 145
25 thisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 42
|||||
146 GCACACCCGGAATACTGTCGCCGACAGGAGCC...AGCGGCATGTGCA 192
42 lYThrValValSerThrThrAsnPhelIleLysAlaGluTyrLysGlyArg 58
|||||
193 CAACGCTCATCTCTCAATGGCTACCTCTCCAGGAGTATTCAGGCAGA 242
59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75
|||||
243 GCCAACCTCATCACTCCAGAGACACACACATTTGTGATTAAACATTGA 292
75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMetA 92
|||||
293 GCAGCTCACCCAGGACGACACTGGGAGCTACAAAGTGTGGCTGGGT...A 339
92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnVal 105
|||||
340 CCAGTAACCGAGCGCTGTCTTCGATGTCAAGCTGGAGGTC 380

```

seq\_name: gb\_ro:MMPIGR3

seq\_documentation\_block:  
LOCUS MMPIGR3 505 bp DNA ROD 09-APR-1998  
DEFINITION Mus musculus pigr gene, exon 3.  
ACCESSION Y16525  
VERSION Y16525.1 GI:3046750  
KEYWORDS pigr gene; polymorphic immunoglobulin receptor.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1 (bases 1 to 505)  
AUTHORS de Groot,N., Vollebregt,E., Lee,S.H., Verbeet,M.P. and de Boer,H.A.  
TITLE Molecular cloning and exon-intron organization of the gene encoding  
the murine polymorphic immunoglobulin receptor  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 505)

AUTHORS de Groot,N.  
TITLE Direct Submission  
JOURNAL Submitted (11-FEB-1998) N. De Groot, Medical Biotechnology Dept.,  
Leiden University, P.O. box 9502, 2300 RA Leiden, NETHERLANDS  
FEATURES Location/Qualifiers  
source  
1. .505  
/organism="Mus musculus"  
/strain="129"  
/db\_xref="taxon:10090"  
/tissue.type="liver"  
77. .421  
/gene="pigr"  
77. .421  
/gene="pigr"  
/label=ex3  
/usedin=Y16524:pig\_cds  
/usedin=Y16524:pig\_mrna  
BASE COUNT 130 a 136 c 127 g 112 t  
ORIGIN

alignment\_scores:  
Quality: 188.00 Length: 97  
Ratio: 2.648 Gaps: 3  
Percent Similarity: 73.196 Percent Identity: 40.206

## alignment\_block:

US-09-135-238B-2\_COPY\_18\_272 x MMPIGR3 ..

Align seg 1/1 to: MMPIGR3 from: 1 to: 505

```

13 GlyGlySerValThrLeuLysCysProLeuProGlu.....Me 25
|||||
130 GCGGACTCTGTTCCATCAGCTGCTACTACCCAGACACCTCTGTCAACCG 179
25 thisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 42
|||||
180 GCACACCCGGAATACTGTCGCCGACAGGAGCC...AGCGGCATGTGCA 226
42 lYThrValValSerThrThrAsnPhelIleLysAlaGluTyrLysGlyArg 58
|||||
227 CAACGCTCATCTCTCAATGGCTACCTCTCCAGGAGTATTCAGGCAGA 276
59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75
|||||
277 GCCAACCTCATCACTCCAGAGACACACACATTTGTGATTAAACATTGA 326
75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMetA 92
|||||
327 GCAGCTCACCCAGGACGACACTGGGAGCTACAAAGTGTGGCTGGGT...A 373
92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnVal 105
|||||
374 CCAGTAACCGAGCGCTGTCTTCGATGTCAAGCTGGAGGTC 414

```

seq\_name: gb\_pat:E15253

seq\_documentation\_block:  
LOCUS E15253 2932 bp DNA PAT 28-JUL-1999  
DEFINITION Mouse pigr receptor gene.  
ACCESSION E15253  
VERSION E15253.1 GI:5709936  
KEYWORDS JP 1998057066-A/1.  
SOURCE Mus sp.  
ORGANISM Mus sp.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1 (bases 1 to 2932)  
AUTHORS Sato,T. and Kushi,A.  
TITLE MURINE PIG RECEPTOR PROTEIN GENE  
JOURNAL Patent: JP 1998057066-A 03-MAR-1998;  
YAKULT HONSHA CO LTD  
COMMENT OS Mus sp. (mouse)



PN JP 1998057066-A/1  
PD 03-MAR-1998  
PF 19-AUG-1996 JP 1996217154  
PI SATO TAKASHI, KUSHIRO AKIRA  
PC C12N15/09,C07H21/04,C07K14/705;  
CC strandedness: Double;  
FH topology: Linear;  
FH key Location/Qualifiers  
FT source 1. .29392 /organism="Mus sp." /clone\_lib="genomic DNA from female 129SVJ FT  
FT mouse liver in  
FT lambda Fix vector'.  
FT Location/Qualifiers  
1. .29392 /organism="Mus sp." /db\_xref="taxon:10095"  
BASE COUNT 8318 a 6747 c 6514 g 7813 t  
ORIGIN  
alignment\_scores:  
Quality: 188.00 Length: 97  
Ratio: 2.648 Gaps: 3  
Percent Similarity: 73.196 Percent Identity: 40.206  
alignment\_block:  
US-09-135-238B-2\_COPY\_18\_272 x E15253 ..  
Align seg 1/1 to: E15253 from: 1 to: 29392  
13 GlyGlySerValThrIleLysCysProLeuProGlu.....Me 25  
||| |||||:|||| ||| |||||:|||||  
18245 GCGACTCTGTTCCATCAGTCTACTACCCAGACACATCTGTCAACCG 18294  
25 tHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 42  
||| ||||| ||| |||||:|||| ||| |||||:|||||  
18295 GCACACCGGAATACTGTGTCGACAGAGGCC...AGCGGCATGTGCA 18341  
42 lyThrValValSerThrAsnPheIleLysAlaGluTyrLysGlyArg 58  
||| |||||:|||||:|||||:||||| |||||:|||||:|||||  
18342 CAAGTCATCTCTTCAATGGCTACTCTCCAAGGAGTATTCAGGCAGA 18391  
59 ValThrLeuLysGlyTyrProArgLysAsnLeuPheLeuValGluValTh 75  
||| |||||:|||||:|||||:||||| |||||:|||||:|||||  
18392 GCCAACCTCATCAACTCCAGACACACACATTTGTGATTAACTGA 18441  
75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 92  
||| |||||:|||||:|||||:||||| ||| ||||| |||  
18442 GCAGCTCACCCAGGAGCAGACTGGGAGCTACAAGTGTGGCTGGGT...A 18498  
92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnVal 105  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
18489 CCAGTAACCGAGCGCTGCTCTTCGATGTCAGCCTGGAGGTC 18529  
seq\_name: gb\_ro:AB001489  
seq\_documentation\_block:  
LOCUS AB001489 29392 bp DNA 14-APR-2000  
DEFINITION Mus musculus DNA for polyimmunoglobulin receptor, complete cds.  
ACCESSION AB001489  
VERSION AB001489.1 GI:2804245  
KEYWORDS polyimmunoglobulin receptor.  
SOURCE Mus musculus (strain:129SVJ) female liver DNA.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1 (sites)  
AUTHORS Kushiro,A. and Sato,T.  
TITLE Polymetric immunoglobulin receptor gene of mouse: sequence,  
structure and chromosomal location  
JOURNAL Gene 204 (1-2), 277-282 (1997)

MEDLINE 98094283  
REFERENCE 2 (bases 1 to 29392)  
AUTHORS Sato,T.  
TITLE Direct Submission  
JOURNAL Submitted (26-FEB-1997) to the DDBJ/EMBL/GenBank databases. Takashi  
Sato, Yakult Central Institute for Microbiological Research,  
Culture Collection Center; 1796 Yaho, Kunitachi-shi, Tokyo 186,  
Japan (Tel:425-77-8960, Fax:425-77-3020)  
Location/Qualifiers  
1. .29392 /organism="Mus musculus"  
/strain="129SVJ"  
/db\_xref="taxon:10090"  
/chromosome="1"  
/map="1E3"  
/sex="female"  
/tissue\_type="liver"  
CAAT\_signal 2470..2474  
TATA\_signal 2580..2585  
CAAT\_signal 2878..2882  
CAAT\_signal 3176..3180  
CAAT\_signal 3251..3255  
TATA\_signal 3350..3355  
5'UTR 3389..3501  
/note="based on the putative transcription start"  
exon 3389..3501  
number=1  
intron 3502..11067  
number=1  
gene 11124..27591  
/gene="pigr"  
join(11124..11166,18192..18536,21130..21783,23022..23354,  
23908..24240,25133..25325,25686..25813,26183..26314,  
26426..26484,27496..27591)  
/gene="pigr"  
/function="mediates secretion of Iga and Igm"  
/codon\_start=1  
/product="polyimmunoglobulin receptor"  
/protein\_id="BAA24431.1"  
/db\_xref="GI:2804246"  
/translation="NRLYLFTLLVTFVSGVSTKSPFGPQEVSSIEGDSVSTCYPD  
TSVNRHTRKYWCRCQAGSMCTTLISSNGYLSKEYSGRANLINFENNTFVINIQLTQ  
DDTGSYKGLGTSNRGLSFDYSLOYPELPSDRHYTKDIGNVTIECPFKRENAP  
SKKSLCKTNSCELYIDSTEKYNESYIGRAKFMKGDLLIVFYVNIHSLTHNDAGLY  
TCDVILNTLGRKDPDFEGRILITPKDDNGRFSVLITGLRKEDAGHYQCAHSSGLPQE  
GWPIQTWOLFVNEESTIPNRRSVVKGTVGGSVAIACYPNPKESSLYKWCWEGDNG  
HCPVLVGTQAOVQEEYEGRLALFDQPGNGTIVTILNOLTTEDAGFYWCLTNGDSWRT  
TTELOVAEATREPNEVTPONATAVLGETFTVSYCHYCKFYSQEKYCKWKNKGCCHIL  
PSHDEGARQSSVSCDQSSQLYSMTLNPVSKDEGWICGVKQGTITGETTAIYIYAVEE  
KTRGSSHVNPTDANARAKVALEEEVDSSI SEKENKAIPNPGFANEREIKNVGDQAO  
ENRSGDAGSADGQSSRSKVLFTLPLGLVLAVGAIAVYAVRHRKNVDRMSIS  
SYRTDISMADFKNSRDGLGNDNMGASPDQTQOTVIEGKDEIVTTTCTAEPEESKAKR  
SSKEADMAYSAFLLQSSSTIAAQVHDGPQEA"  
11124..11166  
/gene="pigr"  
number=2  
11167..18191  
/gene="pigr"  
number=2  
18192..18536  
/gene="pigr"  
number=3  
18537..21129  
/gene="pigr"  
number=3  
21130..21783  
/gene="pigr"  
number=4  
21784..23021  
/gene="pigr"  
number=4  
23022..23354  
exon  
intron  
exon  
intron  
exon  
intron  
exon



Align seg 1/1 to: RNIGR from: 1 to: 3269

```

13 GlyGlySerValThrIleLysCysProLeuProGlu.....Me 25
||||:||||:||||:||||:||||:||||:||||:||||:||||:
170 GGTAACTGGTCTCCATCAGTGTCTACTACCCAGACACTGTGTCAACGG 219
25 thisValArgIleThrLeuCysArgGluMetAlaGlySerGlyThrCysG 42
||||:||||:||||:||||:||||:||||:||||:||||:||||:
220 GCACACCCGGAATACTGGTGCACAGGAGCC...AACGGCTACTGG 266
42 lyThrValValSerThrThrAsnPhelleLysAlaGluThrLysGlyArg 58
||||:||||:||||:||||:||||:||||:||||:||||:||||:
267 CAACCTCATCTCTTCAATAGCTACTCTCGAAGGAGTATTTCAGGCAGA 316
59 ValThrLeuLysGluThrProArgLysAsnLeuPheLeuValGluValTh 75
||||:||||:||||:||||:||||:||||:||||:||||:||||:
317 GCCACCTCATCACTCCAGAGAAATAGCACATTGTGATTAACTATTCG 366
75 rGlnLeuThrGluSerAspSerGlyValThrAlaCysGlyAlaGlyMetA 92
||||:||||:||||:||||:||||:||||:||||:||||:||||:
367 ACATCTCACCCAGGAGGACACTGGGAGCTACAAGTGTGGTCTGGGT...A 413
92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 108
||||:||||:||||:||||:||||:||||:||||:||||:||||:
414 CCATCAACGAGGCGCTGTTTTCGATGTCTCAGCTGGAGGTC..... 454
109 TyrGluProSerTrpGluGluGlnProMetProGluThrProLysTrpPh 125
||||:||||:||||:||||:||||:||||:||||:||||:||||:
455 .....AGCCAGGTCTCTGAGTCCCAAAATGACAC 483
125 eHisLeuProThrLeuPheGlnMetProAlaThrAlaSerSerLysP 142
||||:||||:||||:||||:||||:||||:||||:||||:||||:
484 CCATGTC.....TACACAAAGG 500
142 heValThrArgValThrThrProAlaGlnArgGlyLysValProVal 158
||||:||||:||||:||||:||||:||||:||||:||||:||||:
501 ACATAGGAGAGAACTGTGACCACTCGAATCCGCTTTCAAAGAGGGGAATGCT 550
159 HisHisSer.SerProThrThrGlnIleThrHisArgPro 171
||||:||||:||||:||||:||||:||||:||||:||||:||||:
551 CATAGCAAGAAATCCCTGTGTAAGAAGAGAGAGAGGCGCT 590

```

seq\_name: gb\_om: AB032195

```

seq_documentation_block:
LOCUS      AB032195      2461 bp      mRNA      MAM      03-FEB-2000
DEFINITION Sus scrofa mRNA for poly-Ig receptor, complete cds.
ACCESSION  AB032195
VERSION    AB032195.2 GI:6863079
KEYWORDS   poly-Ig receptor.
SOURCE     Sus scrofa female cDNA to mRNA, clone:SC1.
ORGANISM   Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE  1 (bases 1 to 2461)
AUTHORS    Sone,T. and Kumura,H.
TITLE      Porcine mammary gland cDNA clone, similar to poly-Ig receptor
JOURNAL    Published Only in DataBase (1999) In press
REFERENCE  2 (bases 1 to 2461)
AUTHORS    Sone,T. and Kumura,H.
TITLE      Direct Submission
JOURNAL    Submitted (07-SEP-1999) to the DDBJ/EMBL/GenBank databases. Teruo
            Sone, Hokkaido University, Laboratory of Applied Microbiology,
            Faculty of Agriculture: Kita-9, Nishi-9, Kita-ku, Sapporo, Hokkaido
            060-8589, Japan (E-mail:sonet@chem.agr.hokudai.ac.jp,
            Tel:81-11-706-2502, Fax:81-11-706-4961)
COMMENT    On Feb 3, 2000 this sequence version replaced gi:5881588.
            Sequence updated (01-Feb-2000).
FEATURES   Location/Qualifiers
            1..2461
            /organism="Sus scrofa"
            /db_xref="taxon:9823"
            /clone="SC1"

```

COMMENT

FEATURES

source

```

/sex="female"
sig_peptide 58..111
/product="putative"
CDS         58..2334
            /codon_start=1
            /product="poly-Ig receptor"
            /protein_id="BA084283.2"
            /db_xref="GI:6863080"
            /translation="MTRFFYACLLAIFPVVSMKSPIFGPODVSSVEGSSVSIRCYPA
TSVNRHSRYKRCIGAGKRCCTTLISSEGYISKDYKGRANLTFNPENGTFYMDIGHLTR
GDSLYKRGKIGISSRGLSFDVLEVSQPGQIGDVHYVTADLGSVTVINCPFKSENAQ
KPKSVYKLGQIRVLVDSNGLNNTNRAHLSIOGTNOLVSEFVFNRIQLRDAGY
ICQAGDESSADLQVLKPEPELYIGDLRGSVTFDCALGOEMANYFLCQLKNGKTCN
VYNTLGKKAQDEFEGRILLTPKNSHFSVHITLKRDEAGHYLCGHPDGEPEKGPV
QAWQLFINEDTMIIPRSSVVRVGGVAVTCYPNKPETNSLKYWCWKEENENGRCPO
LVESGLVNEOYERLALYBEPNGTFTVLNQLNTRDAGFYWCLTNEDSRWRSTVEL
KIVGEPNLKLPENVTAWGETLKLSCFFPKFYQKYCKWKNSTGCRALPQDQEQ
SOAFVNDKKSQIISLNLPVRDEGWYCKVDGLHYGETGAVTVAVPQKAKGSD
ARLANAAPAPAEADIEPRARENEVLDPSLIAEDRAVKDGGSPASGSGVPAPOSS
VCGGSGKAVVSTLPLALVNLVGVNTIGVLRHHRKNVDRIISRTRTDLISDFES
SRDGAHDNAGATLDTOETSLGGKDRFTTTTENTMEDKEPKRAKRSKEADKAFITF
LQANSIIAATQNGPREA"
BASE COUNT 608 a 699 c 708 g 446 t
ORIGIN

```

```

alignment_scores:
Quality: 181.00      Length: 118
Ratio: 2.207        Gaps: 4
Percent Similarity: 69.492 Percent Identity: 33.898

alignment_block:
US-09-135-238b-2_copy_18_272 x AB032195 ..

Align seg 1/1 to: AB032195 from: 1 to: 2461

2 IleLeuProGluValLysValGluGlyGluLeu..... 12
||||:||||:||||:||||:||||:||||:||||:||||:||||:
91 ATCTTCCAGTGGTCTCCATGAAGAGTCCCATATTCGTGTCGCCAGGATGT 140
13 .....GlyGlySerValThrIleLysCysProLeuPro.... 23
||||:||||:||||:||||:||||:||||:||||:||||:||||:
141 GAGCAGCGTGGAGGACGCTCGGTGTCCATCAGATGCTACATCCAGCCA 190
24 .....GluMethHisValArgIleThrLeuCysArgGluMetAlaGly 37
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
191 CTTCCGTCACCGCATTCCTCGGAAGTACTGTGTCGCA...ATAGAGCC 237
38 SerGlyThrCysGlyThrValValSerThrThrAsnPhelleLysAlaG1 54
||||:||||:||||:||||:||||:||||:||||:||||:||||:
238 AAGGCCCTGCACACCCCTCATCTCCGAGGGCTACATCTCCAAGGA 287
54 uTyrlsGlyArgValThrLeuLysGlnTyProArgLysAsnLeuPheL 71
||||:||||:||||:||||:||||:||||:||||:||||:||||:
288 CTACAAGGCGAGAGCAACCTCACCACTCCAGAGAACGACCTTCG 337
71 euValGluValThrGlnLeuThrGluSerAspSerGlyValThrAlaCys 87
||||:||||:||||:||||:||||:||||:||||:||||:||||:
338 TGATGGACATTGGCCACCTACCCGCGGTGACTCTGGGCTCTACAAGTGT 387
88 GlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLysValThrLeuAs 104
||||:||||:||||:||||:||||:||||:||||:||||:||||:
388 GGTCTGGCATTAGCAGC...CGAGGCTGTCTTTTGATGTGAGCCTGGA 434
104 nVal 105
|||
435 GGTC 438

```

seq\_name: gb\_pr2:S4343503

seq\_documentation\_block:

```

LOCUS      S4343503      345 bp      DNA      PRI      08-MAY-1993
DEFINITION transmembrane secretory component [human, leukocytes, Genomic, 345

```

nt, segment 3 of 11].  
ACCESSION S43441  
VERSION S43441.1 GI:255089  
KEYWORDS  
SEGMENT 3 of 11  
SOURCE human leukocytes.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 345)  
AUTHORS Krajcic,P., Kvale,D., Tasken,K. and Brandtzaeg,P.  
TITLE Molecular cloning and exon-intron mapping of the gene encoding  
human transmembrane secretory component (the poly-Ig receptor)  
JOURNAL Eur. J. Immunol. 22 (9), 2309-2315 (1992)  
MEDLINE 92387236  
REMARK GenBank staff at the National Library of Medicine created this  
entry [NCBI gibbsq 113232] from the original journal article.  
This sequence comes from Fig. 2.  
FEATURES  
source  
1..345  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 78 a 102 c 97 g 68 t  
ORIGIN  
alignment\_scores  
Quality: 180.50 Length: 106  
Ratio: 2.314 Gaps: 4  
Percent Similarity: 73.585 Percent Identity: 38.679  
alignment\_block  
US-09-135-238B-2\_COPY\_18\_272 x S43435S03 ..  
Align seg 1/1 to: S43435S03 from: 1 to: 345  
4 ProGluValLysValGluGlyGluLeuGlyGlySerValThrIleLysCys 20  
||||| :|||:||||| :|||:|||||:||||| :|||  
30 CCCGAG...GAGGTGAATAGTGTGGAAGGTAACCTCAGTGTCCATCAGCTG 76  
20 sProLeuPro.....GluMetHisValArgIleTyrLeuCysA 33  
||| :|||:||||| :|||:||||| :|||:|||||  
77 CTACTACCCACCCTCTGTCAACCGGCACACCCGGAAGTACTGTGTC 126  
33 rgGluMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsn 49  
||||| :|||:||||| :|||:|||||:||||| :|||  
127 GCGAGGGAGCTAGAGGTGGC...TGCATAACCCCTCATCTCTCGGAGGCG 173  
50 PheIleLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProAr 66  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
174 TAGCTCTCCAGCAATATGTCAGGAGGCTAACCTCACCACACTCCCGA 223  
66 gLysAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerg 83  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
224 GAACGGCACATTGTGGTGAACATTGCCAGCTGAGCGAGGATGATCCG 273  
83 lyValTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGln 99  
||| :|||:||||| :|||:|||||:||||| :|||  
274 GGCCTACAAAGTGTGGCTGGCATCAATAGC...CGAGGCTGTGCTTT 320  
100 LysValThrLeuAsnVal 105  
|||||:|||||:|||||  
321 GATGTCAGCTGGAGGTC 338  
seq\_name: gb\_pat:A52094  
seq\_documentation\_block:  
LOCUS A52094 1936 bp DNA PAT 11-MAR-1997  
DEFINITION Sequence 4 from Patent WO9618734.  
ACCESSION A52094  
VERSION A52094.1 GI:2304706  
KEYWORDS  
SOURCE unidentified.

ORGANISM unidentified  
unclassified.  
REFERENCE 1 (bases 1 to 1936)  
AUTHORS Van,O.J., Asselbergs,F.A., Leist,C. and Gruetter,M.  
TITLE PRODUCTION OF RECOMBINANT SECRETORY COMPONENT  
JOURNAL Patent: WO 9618734-A 4 20-JUN-1996;  
CIBA GEIGY AG (CH)  
COMMENT Other publication AU 4302796 960703.  
FEATURES  
source  
1..1936  
/organism="unidentified"  
/db\_xref="taxon:32644"  
mat\_peptide 16..1920  
/product="TRANSCRIPT FROM SC DNA FRAGMENT 2"  
BASE COUNT 451 a 521 c 583 g 381 t  
ORIGIN  
alignment\_scores  
Quality: 180.50 Length: 106  
Ratio: 2.314 Gaps: 4  
Percent Similarity: 73.585 Percent Identity: 38.679  
alignment\_block  
US-09-135-238B-2\_COPY\_18\_272 x A52094 ..  
Align seg 1/1 to: A52094 from: 1 to: 1936  
4 ProGluValLysValGluGlyGluLeuGlyGlySerValThrIleLysCys 20  
||||| :|||:||||| :|||:|||||:||||| :|||  
88 CCCGAG...GAGGTGAATAGTGTGGAAGGTAACCTCAGTGTCCATCAGCTG 134  
20 sProLeuPro.....GluMetHisValArgIleTyrLeuCysA 33  
||| :|||:||||| :|||:||||| :|||:|||||  
135 CTACTACCCACCCTCTGTCAACCGGCACACCCGGAAGTACTGTGTC 184  
33 rgGluMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsn 49  
||||| :|||:||||| :|||:|||||:||||| :|||  
185 GCGAGGGAGCTAGAGGTGGC...TGCATAACCCCTCATCTCTCGGAGGCG 231  
50 PheIleLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProAr 66  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
232 TAGCTCTCCAGCAATATGTCAGGAGGCTAACCTCACCACACTCCCGA 281  
66 gLysAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerg 83  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
282 GAACGGCACATTGTGGTGAACATTGCCAGCTGAGCGAGGATGATCCG 331  
83 lyValTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGln 99  
||| :|||:||||| :|||:|||||:||||| :|||  
332 GCGCTACAAAGTGTGGCTGGCATCAATAGC...CGAGGCTGTGCTTT 378  
100 LysValThrLeuAsnVal 105  
|||||:|||||:|||||  
379 GATGTCAGCTGGAGGTC 396  
seq\_name: gb\_pat:A52093  
seq\_documentation\_block:  
LOCUS A52093 2031 bp DNA PAT 11-MAR-1997  
DEFINITION Sequence 3 from Patent WO9618734.  
ACCESSION A52093  
VERSION A52093.1 GI:2304705  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified  
unclassified.  
REFERENCE 1 (bases 1 to 2031)  
AUTHORS Van,O.J., Asselbergs,F.A., Leist,C. and Gruetter,M.  
TITLE PRODUCTION OF RECOMBINANT SECRETORY COMPONENT  
JOURNAL Patent: WO 9618734-A 3 20-JUN-1996;  
CIBA GEIGY AG (CH)  
COMMENT Other publication AU 4302796 960703.

```
FEATURES
Source          Location/Qualifiers
1..2031
/organism="unidentified"
/db_xref="taxon:32644"
mat_peptide    111..2015
/product="TRANSCRIPT FROM SC DNA FRAGMENT 1"
BASE COUNT    477 a 540 c 607 g 407 t
ORIGIN

alignment_scores:
  Quality: 180.50      Length: 106
  Ratio: 2.314         Gaps: 4
  Percent Similarity: 73.585      Percent Identity: 38.679

alignment_block:
US-09-135-238B-2_COPY_18_272 x A52093 ..
Align seg 1/1 to: A52093 from: 1 to: 2031

4 ProGluValLysValGluGlyGluLeuGlySerValThrIleIysCy 20
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
183 CCCAG...GAGTGAATAGTGGAGGTAACCTCAGTCTCCATCAGCGT 229

20 sProLeuPro.....GluMetHisValArgIleTyrLeuCysA 33
| ||| :|||:|||||:|||||:|||||:|||||:|||||
230 CTACTACCCACCACCTCTGTCAACCGGCACACCCCGAAGTACTGGTGC 279

33 rgGluMetAlaGlySerGlyThrCysGlyThrValValSerThrAsn 49
||||| ||| :|||:|||||:|||||:|||||:|||||:|||||
280 GGCAGGAGCTAGAGTGGC...TGCATAACCCCTCATCTCCTCGGAGGC 326

50 PheIleLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProAr 66
:|||||:|||||:|||||:|||||:|||||:|||||:|||||
327 TAGCTCTCCAGCAATATGCAGCAGGCTACCTCACCACCTCCCGGA 376

66 gLysAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerG 83
:|||||:|||||:|||||:|||||:|||||:|||||:|||||
377 GAACGGCACATTCGTGGTGAACATTCGCCAGCTGAGCCAGGATGACTCG 426

83 lyValTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGln 99
||| ||| ||||| |||||:|||||:|||||:|||||:|||||
427 GCGCTACAGTGTGCGCTGGCATCAATAGC...CGAGGCGCTGCTCTTT 473

100 LysValThrLeuAsnVal 105
|||||:|||||:|||||
474 GATGTCAGCTGGAGGTC 491

seq_name: gb_pat:A52091

seq_documentation_block:
LOCUS A52091 2405 bp DNA PAT 11-MAR-1997
DEFINITION Sequence 1 from Patent WO9618734.
ACCESSION A52091
VERSION A52091.1 GI:2304703
KEYWORDS .
SOURCE .
ORGANISM unidentified.
unclassified.
REFERENCE
1 (bases 1 to 2405)
AUTHORS Van,O.J., Asselbergs,F.A., Leist,C. and Gruetter,M.
TITLE PRODUCTION OF RECOMBINANT SECRETORY COMPONENT
JOURNAL Patent: WO 9618734-A 1 20-JUN-1996;
CIBA GEIGY AG (CH)
COMMENT Other publication AU 4302796 960703.
FEATURES
source
Location/Qualifiers
1..2405
/organism="unidentified"
/db_xref="taxon:32644"
111..2405
/codon_start=1
/product="ORIGINAL PIGR"
/protein_id="CAA03384.1"
```



OM of: US-09-135-238B-2\_COPY\_18\_272 to: Issued\_Patents\_NA.\* out\_format : pfs  
Date: Sep 12, 2000 6:53 AM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:  
-MODEL=framet\_p2n.model -DBV=xlp  
-Q/cgn2\_1/USPTO\_spool/US09135238/runat\_24082000\_091324\_17895/app\_query.fasta\_1.1282  
-DB=Issued\_Patents\_NA -QFMT=fastap -SUFFIX=p2n.rni  
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000  
-LOAPEXT=0.000 -OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000  
-XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000  
-YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1  
-MATRIX=blosom62 -TRANS=human40.cdi -LIST=45 -DOCLALIGN=200  
-THRCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=1000000  
-USER=US09135238\_@CGN1\_1\_68 -NCPU=6 -ICPU=3 -LONGLOG -NO\_XLPXY  
-WAIT -THREADS=1

Search information block:  
Query: US-09-135-238B-2\_COPY\_18\_272  
Query length: 255  
Database: Issued\_Patents\_NA.\*  
Database sequences: 243050  
Database length: 68777915  
Search time (sec): 93.130000

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/cgn2_6/ptodata/1/ina/6/COMB.seq:US-08-434-000A-7	+	197.00	340.00	1.9e-11	3095
/cgn2_6/ptodata/1/ina/6/COMB.seq:US-08-434-000A-9	+	185.50	317.08	3.6e-10	3269
/cgn2_6/ptodata/1/ina/6/COMB.seq:US-08-434-000A-3	+	180.50	308.80	1.0e-09	2919
/cgn2_6/ptodata/1/ina/6/COMB.seq:US-08-434-000A-5	+	179.50	304.19	1.9e-09	3630
/cgn2_6/ptodata/1/ina/5D/COMB.seq:US-08-642-406A-21	+	166.50	287.15	1.7e-08	1875
/cgn2_6/ptodata/1/ina/5D/COMB.seq:US-08-642-406A-20	+	166.50	279.42	4.5e-08	3517
/cgn2_6/ptodata/1/ina/6/COMB.seq:US-08-434-000A-1	+	166.50	279.42	4.5e-08	3517
/cgn2_6/ptodata/1/ina/5A/COMB.seq:US-08-286-740-3	+	102.00	146.97	1.07	6557
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US95-09576-3	+	102.00	146.97	1.07	6557
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-378-939-9	+	100.50	161.27	0.1716	1617
/cgn2_6/ptodata/1/ina/5D/COMB.seq:US-08-887-352B-1	+	99.50	142.97	1.79	6127
/cgn2_6/ptodata/1/ina/5D/COMB.seq:US-08-935-450-10	+	99.00	185.38	0.1013	914
/cgn2_6/ptodata/1/ina/5A/COMB.seq:US-07-841-646-3	-	97.00	125.30	17.29	17410
/cgn2_6/ptodata/1/ina/5B/COMB.seq:US-08-147-023-3	-	97.00	125.30	17.29	17410
/cgn2_6/ptodata/1/ina/5B/COMB.seq:US-08-447-570-3	-	97.00	125.30	17.29	17410
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-449-700-3	-	97.00	125.30	17.29	17410
/cgn2_6/ptodata/1/ina/5D/COMB.seq:US-08-449-699A-3	-	97.00	125.30	17.29	17410
/cgn2_6/ptodata/1/ina/5B/COMB.seq:US-08-157-101A-9	+	96.50	155.75	0.3484	1350
/cgn2_6/ptodata/1/ina/5B/COMB.seq:US-08-157-101A-6	+	96.50	153.85	0.4446	1576
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-932-376A-1	+	93.50	146.84	1.09	1738
/cgn2_6/ptodata/1/ina/5B/COMB.seq:US-08-463-587A-24	+	92.00	141.17	2.26	2178
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-463-667A-2	+	92.00	141.17	2.26	2178
/cgn2_6/ptodata/1/ina/6/COMB.seq:US-08-923-854-24	+	92.00	141.17	2.26	2178
/cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:PCT-US91-09133-25	+	92.00	141.17	2.26	2178
/cgn2_6/ptodata/1/ina/5B/COMB.seq:US-08-149-099C-3	+	91.50	122.49	24.82	9209
/cgn2_6/ptodata/1/ina/5B/COMB.seq:US-08-476-275-2	+	91.50	122.49	24.82	9209
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-478-967A-3	+	91.50	122.49	24.82	9209
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-819-866-2	+	91.50	113.60	77.61	18986
/cgn2_6/ptodata/1/ina/5D/COMB.seq:US-09-023-715-2	+	91.50	113.60	77.61	18986
/cgn2_6/ptodata/1/ina/5B/COMB.seq:US-08-592-126-75	-	91.00	147.81	0.9637	1083
/cgn2_6/ptodata/1/ina/5C/COMB.seq:US-08-932-376A-3	+	91.00	139.00	2.99	2220
/cgn2_6/ptodata/1/ina/6/COMB.seq:US-09-027-449-59	+	90.50	148.76	0.8539	927
/cgn2_6/ptodata/1/ina/6/COMB.seq:US-09-027-449-61	+	90.50	124.71	18.65	6563
/cgn2_6/ptodata/1/ina/6/COMB.seq:US-09-027-449-68	+	90.50	122.10	26.09	8120
/cgn2_6/ptodata/1/ina/6/COMB.seq:US-09-027-449-68	+	90.50	119.57	36.06	9972
/cgn2_6/ptodata/1/ina/5A/COMB.seq:US-08-276-099A-1	+	90.00	133.18	6.30	3046
/cgn2_6/ptodata/1/ina/5B/COMB.seq:US-08-781-890-1	+	90.00	133.18	6.30	3046
/cgn2_6/ptodata/1/ina/5D/COMB.seq:US-08-539-205A-5	+	90.00	132.47	6.90	3226
/cgn2_6/ptodata/1/ina/5B/COMB.seq:US-08-860-882A-28	+	89.50	148.99	0.8287	777
/cgn2_6/ptodata/1/ina/5B/COMB.seq:US-08-394-326-1	+	89.00	135.99	4.39	2070

/cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq:PCT-US93-00031-18 + 88.50 135.89 4.45  
/cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq:PCT-US93-00031-20 + 88.50 135.87 4.46  
/cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq:PCT-US93-00031-10 + 88.50 135.81 4.49  
/cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq:PCT-US93-00031-22 + 88.50 135.81 4.49  
/cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq:PCT-US93-00031-12 + 88.50 134.24 5.49  
seq\_name: /cgn2\_6/ptodata/1/ina/6\_COMB.seq:US-08-434-000A-7  
seq\_documentation\_block:  
; Sequence 7, Application US/08434000A  
; Patent No. 6046037  
; GENERAL INFORMATION:  
; APPLICANT: ANDREW C. HIATT, JULIAN  
; APPLICANT: K.-C. MA, THOMAS LEHNER  
; TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LYON & LYON  
; STREET: 633 West Fifth Street  
; STREET: Suite 4700  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 90071  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; MEDIUM TYPE: Storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: IBM P.C. DOS 5.0  
; SOFTWARE: Word Perfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/434,000A  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA: including application  
; PRIOR APPLICATION DATA: described below: 1  
; APPLICATION NUMBER: 08/367,395  
; FILING DATE: 12/30/94  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Guise, Jeffrey W.  
; REGISTRATION NUMBER: 34,613  
; REFERENCE/DOCKET NUMBER: 212/127  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 552-8400  
; TELEFAX: (619) 552-0159  
; TELEX: 67-3510  
; SEQUENCE LISTING  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3095 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; TOPOLOGY: DESCRIPTION: Mouse Polyimmunoglobulin Receptor  
; FEATURE:  
; NAME/KEY: Coding Sequence  
; LOCATION: 85....2400  
; US-08-434-000A-7  
alignment\_scores:  
Quality: 197.00 Length: 148  
Ratio: 2.165 Gaps: 5  
Percent Similarity: 61.486 Percent Identity: 32.432  
alignment\_block:  
US-09-135-238B-2\_COPY\_18\_272 x US-08-434-000A-7 ..  
Align seq 1/1 to: US-08-434-000A-7 from: 1 to: 3095  
13 GlycylSerValThrIleLysCysProLeuProGlu.....Me 25

```
181 GGGGACTCTGTTTCCATCAGCTGCTACTACCCAGACACCTCTGTCAACCG 230
25 tHisValArgIleThrLeuCysArgGluMetAlaGlySerGlyThrCysG 42
231 GCACACCCGGAATACGTGGCCGCAAGGAGCC...AGCGCATGTGCA 277
42 lyThrValValSerThrThrAsnPheIleLysAlaGluThrLysGlyArg 58
278 CAACGCTCATCTCTCAATGCTACTCTCCAGGAGATTTCAGGCAGA 327
59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75
328 GCAACCTCATCACTTCCAGAGAACACACATTTGTGATTAACATTGA 377
75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 92
378 GCAGCTCACCCAGCAGCACACTGGGAGCTACAAGTGGCGTGGCT...A 424
92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 108
425 CCAGTAACCGAGGCTGCTCTCGATGTCAGCCTGGAGGTC..... 465
109 TyrGluProSerTrpGluGluGlnProMetProGluThrProLysTrpPh 125
466 .....AGCCAGGTCTCTGAGTGGCGAGTGACAC 494
125 eHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSerSerLysP 142
495 CCAGCTC.....TACAAAGGACATAGGCA 520
142 heValThrArgValThrProAlaGlnArgGlyLysValPro 156
521 GAATGTGACCATGATGCTCTTCAAAAGGGAGATGTCCC 564

seq_name: /cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-434-000A-9
seq_documentation_block:
; Sequence 9, Application US/08434000A
; Patent No. 6046037
; GENERAL INFORMATION:
; APPLICANT: ANDREW C. HIATT, JULIAN
; APPLICANT: K. C. MA, THOMAS LEHNER
; TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
; TITLE OF INVENTION: PROTEINS IN PLANTS AND THEIR USES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/434, 000A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below:
; APPLICATION NUMBER: 08/367,395
; FILING DATE: 12/30/94
; ATTORNEY/AGENT INFORMATION:
; NAME: Guise, Jeffrey W.
; REGISTRATION NUMBER: 34,613
; REFERENCE/DOCKET NUMBER: 212/127
```

```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 552-8400
; TELEFAX: (619) 552-0159
; TELEX: 67-3510
; TELETYPE: SEQUENCE LISTING
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3269 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; TOPOLOGY: DESCRIPTION: Rat Polyimmunoglobulin Receptor
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 74....2383
US-08-434-000A-9

alignment_scores:
Quality: 185.50 Length: 164
Ratio: 1.855 Gaps: 6
Percent Similarity: 60.976 Percent Identity: 30.488

alignment_block:
US-09-135-238B-2_copy_18_272 x US-08-434-000A-9 ..
Align seg 1/1 to: US-08-434-000A-9 from: 1 to: 3269
13 GlyGlySerValThrIleLysCysProLeuProGlu.....Me 25
170 GGTAACTCGGTCTCCATCAGTCGTACTACCCAGACACCTCTGTCAACCG 219
25 tHisValArgIleThrLeuCysArgGluMetAlaGlySerGlyThrCysG 42
220 GCACACCCGGAATACGTGGCCGCAAGGAGCC...AACGGCTACTGCG 266
42 lyThrValValSerThrThrAsnPheIleLysAlaGluThrLysGlyArg 58
267 CAACCTCATCTCTCAAAATGGCTACCTCTCGAAGGAGTATTTCAGGCAGA 316
59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75
317 GCCAGCTCATCACTTCCAGAGAACACACATTTGTGATTAACATTGC 366
75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 92
367 ACATCTCACCCAGGAGCACACTGGGAGCTACAAGTGTGCTGGGT...A 413
92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 108
414 CCACATACCCGAGGCTGTTTTCGATGTCAGCTGGAGGTC..... 454
109 TyrGluProSerTrpGluGluGlnProMetProGluThrProLysTrpPh 125
455 .....AGCCAGGTCTCTGAGTGGAGTTCACCAAGGAGGCAAC 483
125 eHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSerSerLysP 142
484 CCATGTC.....TACACAAAGG 500
142 heValThrArgValThrProAlaGlnArgGlyLysValProProVal 158
501 ACATAGGCAGAACTGTGACCATCGAATGCCGTTTCAAGAGGGAATGCT 550
159 HisHisSer_SerProThrThrGlnIleThrHisArgPro 171
551 TACACAGAAATCCCTGTGTAAGAAGAGAGAGGAGGCT 590

seq_name: /cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-434-000A-3
seq_documentation_block:
; Sequence 3, Application US/08434000A
; Patent No. 6046037
```



```

; GENERAL INFORMATION:
; APPLICANT: ANDREW C. HIATT, JULIAN
; APPLICANT: K.-C. MA, THOMAS LEHNER
; TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
; TITLE OF INVENTION: PROTEINS IN PLANTS AND THEIR USES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Suite 4700
; STATE: Los Angeles
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/434,000A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below: 1
; APPLICATION NUMBER: 08/367,395
; FILING DATE: 12/30/94
; ATTORNEY/AGENT INFORMATION:
; NAME: Guise, Jeffrey W.
; REGISTRATION NUMBER: 34,613
; REFERENCE/DOCKET NUMBER: 212/127
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 552-8400
; TELEFAX: (619) 552-0159
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2919 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; TOPOLOGY: DESCRIPTION: Human polyimmunoglobulin Receptor
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 235....2472
; US-08-434-000A-3

alignment_scores:
  Quality: 180.50      Length: 106
  Ratio: 2.314        Gaps: 4
  Percent Similarity: 73.585      Percent Identity: 38.679

alignment_block:
US-09-135-238b-2_COPY_18_272 x US-08-434-000A-3 ..
Align seg 1/1 to: US-08-434-000A-3 from: 1 to: 2919

      4 ProGluValLysValGluGlyGluLeuLeuGlyGlySerValThrIleLysCy 20
      ||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
253 CCCGAG...GAGTGAATAGTGGAGGTAACCTACGTCTCCATCAGCGT 299
      |
20 sProLeuPro.....GluMethHisValArgIleTyrLeuCy 33
      ||||| :|||:|||||:|||||:|||||:|||||:|||||:
300 CTACTACCCACCCCTCTGTCAACCGGCACACCCGGGAAGTACTGTGTC 349
      |
33 rGluMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsn 49
      ||||| :|||:|||||:|||||:|||||:|||||:|||||:
350 GGCAGGGAGCTAGAGTGGC...TGCATAACCCCTCATCTCTCGGAGGC 396

```



```

; CITY: La Jolla
; STATE: California
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642,406A
; FILING DATE: 03-MAY-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/591,823
; FILING DATE: 02-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/427,765
; FILING DATE: 27-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Logan, April C.
; REGISTRATION NUMBER: 33,950
; REFERENCE/DOCKET NUMBER: 184.2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 554-2937
; TELEFAX: (619) 554-6312
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3517 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-642-406A-20

alignment_scores:
  Quality: 166.50      Length: 103
  Ratio: 2.602        Gaps: 4
  Percent Similarity: 62.136      Percent Identity: 42.718

alignment_block:
US-09-135-238B-2_COPY_18_272 x US-08-642-406A-20  ..
Align seg 1/1 to: US-08-642-406A-20 from: 1 to: 3517

13 GlyGlySerValThrIleLysCysProLeuPro.....GluMe 25
||| |||||:|||| ||| |||
238 GCGACTCGGTGCCATCATGCTACTACCAACAACCTCGTCACCGC 287

25 ThisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 42
||| ||| :||| ||||| ||| ||||| |||
288 GCACACCGGAGTGTGTGCGCGGAA...GAGGAGCGCGCGCTGCG 334
|||:||||| |||:||||| :|||:|||||
42 IyThrValValSerThrThrAsnPhelIeLysAlaGluTyrLysGlyArg 58
||||:||||| |||:||||| :|||:|||||
335 TGACGCTTGCTCG...ACCGCTACAGCTCCCGAGTAATCTCCGGAGA 381

59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75
||| :||||| ||| ||||| ||||| |||
382 GCAACTCACCGACTCCCTCCATAAAGGGAGTTTGTGTGACTGTGA 431

75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMetA 92
|||||:|||||:|||||:||||| ||| |||||:|||||:|||||
432 CCAACTCACCAAGACGACTCAGGAGCTCAAGTGTGGCGTGGAGTCA 481

92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 108
|| ||||| |||:||||| |||:|||||
482 AC...GGCGGTGGCTGGACTTCGGTGTCAACGCTGCTGTCAGCCAGAG 528

109 TyrGluPro lll
```

```

529 CCAGAGCGCT 537
|||||
seq.name: /cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-434-000A-1

seq_documentation_block:
; Sequence 1, Application US/08434000A
; Patent No. 6046037
; GENERAL INFORMATION:
; APPLICANT: ANDREW C. HIATT, JULIAN
; APPLICANT: K.-C. MA, THOMAS LEHNER
; TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
; TITLE OF INVENTION: PROTEINS IN PLANTS AND THEIR USES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: Storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/434,000A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below: 1
; APPLICATION NUMBER: 08/367,395
; FILING DATE: 12/30/94
; ATTORNEY/AGENT INFORMATION:
; NAME: Guise, Jeffrey W.
; REGISTRATION NUMBER: 34,613
; REFERENCE/DOCKET NUMBER: 212/127
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 552-8400
; TELEFAX: (619) 552-0159
; TELEX: 67-3510
; TELEX: SEQUENCE LISTING
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3517 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; TOPOLOGY: DESCRIPTION: Rabbit polyimmunoglobulin receptor
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 124....2445
; US-08-434-000A-1

alignment_scores:
  Quality: 166.50      Length: 103
  Ratio: 2.602        Gaps: 4
  Percent Similarity: 62.136      Percent Identity: 42.718

alignment_block:
US-09-135-238B-2_COPY_18_272 x US-08-434-000A-1  ..
Align seg 1/1 to: US-08-434-000A-1 from: 1 to: 3517

13 GlyGlySerValThrIleLysCysProLeuPro.....GluMe 25
||| |||||:|||| ||| |||
238 GCGACTCGGTGCCATCATGCTACTACCAACAACCTCGTCACCGC 287
```

```
25 tHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 42 *
   ||| ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| |||
288 GCACAGCGGGAAGTCTGTGGCGGAA...GAGGAGAGCGCGCGCTGCG 334
42 lYThrValValSerThrThrAsnPhelIleLysAlaGluTyrLysGlyArg 58
   ||| ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| |||
335 TGACGCTTGCCTCG...ACCGGTACACGTCACGAGTAATCTCCGGAGA 381
59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75
   ||| ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| |||
382 GGCAGGCTCACCGACTCCCTGATAAAGGGGAGTTTGTGCTGACTGTGA 431
75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMetA 92
   ||| ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| |||
432 CCAACTCACCCAGAACGACTCAGGGAGCTACAAGTGTGGCGTGGAGTCA 481
92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 108
   ||| ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| |||
482 AC...GGCGGTGGCTGGACTTCGGTGTCAAGTGTGCTCAGCCAGAAG 528
109 TyrGluPro 111
   ||| |||
529 CCAGAGCGCT 537
```

seq\_name: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:us-08-286-740-3

```
seq documentation block:
; Sequence 3, Application US/08286740
; Patent No. 5581053
; GENERAL INFORMATION:
; APPLICANT: Crowley, Craig W.
; TITLE OF INVENTION: METHOD FOR SELECTING HIGH-EXPRESSING
; TITLE OF INVENTION: HOST CELLS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/286,740
; FILING DATE: 05-AUG-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee, Wendy M.
; REGISTRATION NUMBER:
; REFERENCE/DOCKET NUMBER: 798
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1994
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6557 bases
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-286-740-3
```

alignment\_scores:  
Quality: 102.00  
Ratio: 0.843

Length: 252  
Gaps: 15

```
Percent Similarity: 48.016 Percent Identity: 24.603
alignment_block:
US-09-135-238B-2_COPY_18_272 x US-08-286-740-3 ..
Align seg 1/1 to: US-08-286-740-3 from: 1 to: 6557

13 GlyGlySerValThrIleLysCysProLeuProGluMetHisValArgI 29
   ||| ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1405 GGGGGCTCACTCGGTTTGTCTGTGAGTTTCTGGCTACTCCATCACCTC 1454
29 e.....TyrLeuCysArgGluMetAlaGlySerGly..... 39
   ||| ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1455 CGGATATAGCTGGAAGTGGATCGCTCAGCGCCCGGTAAGGGCTGGAAT 1504
40 .....ThrCysGlyThrValValSerThrThrAsnPhelIleLysAlaGlu 54
   ||| ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1505 GGGTTGCATCGATTACGTATGCCGGATCGACTAAGTAT...AACCCTAGC 1551
55 TyrLysGlyArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLe 71
   ||| ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1552 GTCAGGGCGGTATCACTATTAAGTCGGAGATTCCAAAACACATTCATA 1601
71 uValGluValThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysG 88
   ||| ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1602 CCGTGCAGATGAACAGCGCTGCTGAGGACACACTGCCGCTCTATTATTGTG 1651
88 lYAlaGlyMetAsn...ThrAspArgGlyLysThrGlnLysValThrLeuAs 104
   ||| ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1652 CTCGAGGACGACCATTTTCGGCGCTGCGACTTCGCGTGTGGGGTCAA 1701
104 nValHisSerGluTyrGluProSerTrpGluGlnPrometPro.... 119
   ||| ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1702 G.....GAAACC...TGGTCACCGTCTCCTCGCGCTCCAC 1733
120 .....GluThrProLysTrpPhe.....HisLeuProTyr 129
   ||| ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1734 CAAGGGCCCATGGCTCTCCCGCTGGCACCCCTCTCCAGAGACACCTCTG 1783
130 LeuPheGlnMetProAlaTyrAlaSerSerSerLysPheValThrArgVa 146
   ||| ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1784 GGGGCACAGCGCGCTGGCTGCTGCTGCA.....AGGAC 1818
146 lThrThrProAlaGln.....ArgGlyLysValProProValHisH 160
   ||| ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1819 TACTTCCCCGACCGGTCACGGTCTGCTGGAACCTCAGCGCCCTGACCCAG 1868
160 lSerSerProThrThrGlnIleThrHisArgProArgValSerArgAla 176
   ||| ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1869 CGGGGTGCACACCTTCGGGGCTGTCTTACAGTCTCTCAGGACTCTACTCC 1918
177 SerSer.....ValAlaGlyAspLysProArgThrPh 187
   ||| ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1919 TCAGCAGCGGTGGTACTGTGCCCTTAGCAGCTTGGCACCCAGA..... 1963
187 eleuProSerThrThrAlaSerLysIleSerAlaLeuGluGlyLeuLeuL 204
   ||| ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1964 .....CCTACATCTGCAACGCTGAATCACAAGC..... 1990
204 ysProGlnThrPro.....SerTyrAsnHisHisThr 214
   ||| ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1991 ..CCAGCAACACCAAGGTGGGACAAAGTTGAGCCCAATCTTGTGACACA 2038
215 ArgLeu...HisArgGlnArgAlaLeuAspTyrGlySerGlnSerGlyAr 230
   ||| ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2039 AAACCTCACATATGCCACCGGTGCCAGCACCTGAACTCCTGGGGGGACCG 2088
230 gGlu 231
   ||| |||
2089 TCAG 2092
```

seq\_name: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq:PCT-US95-09576-3





```
Ratio: 0.850 Gaps: 12
Percent Similarity: 45.174 Percent Identity: 23.552
alignment_block:
US-09-135-238B-2_COPY_18_272 x US-08-887-352B-1 ..
Align seg 1/1 to: US-08-887-352B-1 from: 1 to: 6127

13 GlyGlySerValThrIleLysCysProLeuProGluMethHisValArgII 29
|||||.....|
1357 GGGGGCTCACTCCGTTTCTGTGCGAGTTTCTGGCTACTCCATCACTC 1406
29 e.....TyrLeuCysArgGluMetAlaGlySerGly..... 39
1407 CGGATACAGCTGGAACCTGGATCGCTCAGCGCCCGGTAAGGGCTGAAT 1456
40 .....ThrCysGlyThrValValSerThrThrAsnPheIleLysAlaGlu 54
|||||.....|
1457 GGGTTGATCGAATTACGTATGACGGATCGACTAATAT...AACCCTAGC 1503
55 TyrLysGlyArgValThrLysGlnTyrProArgLysAsnLeuPheLe 71
|||||.....|
1504 GTCAGGGCGGTATCATATAGTCGGGACGATTCACAAAACACATTCTA 1553
71 uValGluValThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysG 88
|||||.....|
1554 CTTGCAGATGAACAGCCTCGTCTGAGGACACTGCGCTATATTATTG 1603
88 lYalaGlyMetAsnThrAspArgGlyLys..ThrGlnLysValThrLeuAs 104
|||||.....|
1604 CT.....CGAGGAGCCACTATTTCGTCTACTGGCAC 1635
104 nValHisSerGluTyrGluProSerTrpGluGlnProMetProGluT 121
|||||.....|
1636 TTCGCCGTGGGGTCAAGGAACCTGG.....TCACCGCTCTCCTCGGC 1679
121 hrProLysTrpPheHis.....Leu 127
|||||
1680 CTCACCAAGGGCCCTCGTCTTCCCTTAGCACCTCTCTCCACAGCA 1729
128 ProTyrLeuPheGlnMetProAlaTyrAlaSerSerSerLysPheValTh 144
|||||
1730 CCTCTGGGGGCACAGCGGCCCTGGGTGCTTGCTCA..... 1765
144 rArgValThrProAlaGln.....ArgGlyLysValProProV 158
|||||
1766 AGGACTACTCTCCCGACCGGTGAGCGTGTGTTGAACCTAGGCGCCT 1814
158 alHisHisSerSerProThrThrGlnIleThrHisArgProArgValSer 174
|||||
1815 GACCAGGGGGTGCACACCTTCCGGGTGCTTACAGTCTCTCAGGACTCT 1864
175 ArgAlaSerSerValAlaGlyAspLysProArgThrPheLeuProSerTh 191
|||||
1865 ACTCCCTCAGCAGCGGTGACCGTGCCT.....CCAGCAGC 1902
191 rThrAlaSerLysIleSerAlaLeuGluGlyLeuLysProGlnThrP 208
|||||
1903 TTGGGCACCCAGCAGTACATCTGCAAGTGAATCAACGCCAGCAACAC 1952
208 ro.....SerTyrAsnHisHisThrArgLeuHisArg 218
|||||
1953 CAAGGTGGACAGAAAGTTGAGCCCAATCTTGTGACAAAACCTACACCT 2002
219 GlnArgAlaLeuAspTyrGlySerGlnSerGlyArgGluGlyGlnGlyPh 235
|||||
2003 AGAGTGGCGGTG..... 2014
235 eHisIleLeuIleProThrIleLeu 243
|||||
2015 ....GCTCTGGTTCCGGTGATTGTG 2035

seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-935-450-10
```

```
seq_documentation_block:
; Sequence 10, Application US/08935450
; Patent No. 5977311
; GENERAL INFORMATION:
; APPLICANT: Nandabalan, Krishnan
; APPLICANT: Yang, Meljia
; APPLICANT: Schulz, Vincent
; TITLE OF INVENTION: 53BP2 COMPLEXES
; FILE REFERENCE: 7934-054
; CURRENT APPLICATION NUMBER: US/08/935,450
; CURRENT FILING DATE: 1997-09-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 10
; LENGTH: 914
; TYPE: DNA
; ORGANISM: Homo sapiens
US-08-935-450-10

alignment_scores:
Quality: 99.00 Length: 146
Ratio: 1.456 Gaps: 9
Percent Similarity: 46.575 Percent Identity: 29.452

alignment_block:
US-09-135-238B-2_COPY_18_272 x US-08-935-450-10 ..
Align seg 1/1 to: US-08-935-450-10 from: 1 to: 914

116 GlnProMetProGluThrProLysTrpPheHisLeu..... 127
|||||
232 CAGCCCTGCTCGGAACCCCGGAGGCGGCGAGCACCTACAATAAGAACAGCA 281
128 ....ProTyrLeu.....PheGlnMetProAlaTyrAlaSerSer 140
|||||
282 ACATCCCTGGCTCAAGCCCAATACAGACCCCGCCCGCTCCACCTCCACCCAGCTAC 331
140 erLysPheValThrArgValThrThr.....ProAlaGlnArg..... 152
:: ||| ::|||
332 AGCCCTTCCACAGCGGAGTTACAGCGACGCCACCTACACCCAGGGGAGGT 381
153 .....GlyLysValProPro.ValHisHisSerSerP 163
|||||
382 TACAGCCAGGGTTTACACAGGCCCGCCCGCTCCACCTCCACCCAGCTCCG 431
163 roThrThrGln.....IleThrHisArgProArgValSerArg 175
|||||
432 CTACAATATGGGAGCTACGGCGGTTACACCCCGCCCGCTTATACCCAC 481
176 AlaSerSerValAlaGlyAspLysProArgThrPheLeuProSerThrTh 192
|||||
482 CGCCACCC....CCACCGCACAGACCTACCTCAGCCCAACTATAACAG 528
192 rAlaSerLysIleSerAlaLeuGluGlyLeuLys.....PROG 206
:: ||| ::|||
529 TATCAGCAGTATGCCAGCGTGGAAACCACTACTATCAGAACCCAGGGCCAG 578
206 lnrProSerTyrAsnHisHisThrArgLeuHisArgGlnArgAlaLeu 222
|||||
579 TGGCGCCATACTACGGGAACACTAC..... 601
223 AspTyrGlySerGlnSerGlyArgGluGlyGlnGly 234
|||||
602 GACTACGGGAGTACTCCGGGAACACACAGGGTGGC 637

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-841-646-3

seq_documentation_block:
; Sequence 3, Application US/07841646
; Patent No. 5266683
; GENERAL INFORMATION:
```











a

OM of: US-09-135-238B-2\_COPY\_18\_272 to: EST: \* out\_format : pfs

Date: Sep 12, 2000 6:04 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp  
-O/cgn2.1/USPTO\_spool/US09135238/runat\_24082000\_091323\_17869/app\_query.fasta\_1.1282  
-DB=EST -QFWT=fastap -SUFFIX=p2n.rst -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000 -GAPOP=4.500  
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTFWT=pfs -NORM=ext -MINLEN=0  
-MAXLEN=100000 -USER=US09135238 -CGN1\_1\_2178 -NCPU=6 -ICPU=3  
-LONGLOG -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-135-238B-2\_COPY\_18\_272

Query length: 255

Database: EST: \*

Database sequences: 5247842

Database length: ~2090053206

Search time (sec): 1563.530000

score\_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
gb_est42:AW402953	+	544.00	919.78	2.1e-42	312	! AW402953 UI-HF-BK0-abd-c-06-0-0
gb_est44:AW655210	+	451.00	756.31	2.7e-33	518	! AW655210 105863 MARC 1BOV Bos t
gb_est42:AW444691	-	444.50	746.05	1.0e-32	482	! AW444691 UI-H-B13-ajw-g-12-0-0
gb_est8:AA521993	+	425.00	710.07	1.0e-30	627	! AA521993 vntel2.r1 Soares_mamm
gb_est41:AW327067	+	415.00	699.13	4.1e-30	359	! AW327067 20516 MARC 1BOV Bos t
gb_est45:AW655218	+	384.00	643.01	5.5e-27	494	! AW655218 105873 MARC 1BOV Bos t
gb_est25:AI760373	-	328.00	543.94	1.8e-21	711	! AI760373 wh87d11.x1 NCI CGAP Cl
gb_est12:AA830944	-	305.00	509.01	1.6e-19	485	! AA830944 oc58b10.s1 NCI CGAP Cl
gb_est8:AA491123	+	299.00	509.61	1.5e-19	179	! AA491123 aa46g04.r1 NCI CGAP Cl
gb_est40:AW291041	-	292.50	486.20	3.0e-18	561	! AW291041 UI-H-B12-agj-h-06-0-0
gb_est4:AA289677	+	281.00	469.88	2.4e-17	417	! AA289677 vb20c07.r1 Soares_mous
gb_est3:AA174968	+	218.00	361.79	2.5e-11	460	! AA174968 ms78h03.r1 Soares_mous
gb_est36:AA509857	+	209.50	356.43	5.0e-11	199	! AA509857 vh51a10.r1 Soares_mamm
gb_est3:AA012833	+	180.00	300.94	6.2e-08	327	! AA012833 uc03d11.y1 Sugano mous
gb_est15:AI021484	+	157.50	267.01	4.8e-06	220	! AI021484 ub08h03.r1 Soares_mamm
gb_est3:AA177660	+	145.00	245.97	7.2e-05	216	! AA177660 mt20d11.r1 Soares_mous
gb_est44:AW603866	+	127.00	204.08	0.0154	614	! AW603866 CM0-CN0039-280100-168
gb_gss15:AZ033800	+	121.00	196.09	0.0429	501	! AZ033800 RPCI-23-248K2.TV RPCI
gb_est48:FI4851	+	120.50	200.58	0.0241	306	! FI4851 SSC24F11 Porcine small
gb_est44:AW603303	+	120.50	193.04	0.0635	614	! AW603303 RCI-CN0012-170100-011
gb_est49:HB4070	+	117.00	190.93	0.0833	431	! HB4070 ys4c09.r1 Soares retina
gb_est43:AW484527	+	107.50	175.23	0.6238	414	! AW484527 63153 MARC 3BOV Bos t
gb_est12:AA839051	+	105.50	170.03	1.22	489	! AA839051 vw47a02.r1 Soares_mamm
gb_est49:HB4076	+	105.00	171.06	1.06	411	! HB4076 ys4c12.r1 Soares retina
gb_est17:AI177514	+	102.50	164.20	2.57	523	! AI177514 EST221146 Normalized
gb_est47:CI88490	+	102.50	160.89	3.92	710	! CI88490 C88490 Mouse early blas
gb_est39:AW176541	+	101.50	167.89	1.60	318	! AW176541 MR0-CT0062-200899-002
gb_est40:AW258498	+	101.50	162.50	3.19	523	! AW258498 uq35h10.y1 NCI CGAP Ma
gb_est43:AA127111	+	100.50	161.82	3.48	476	! AA127111 mw81f05.r1 Soares mous
gb_gss8:AQ349923	+	100.50	161.51	3.62	490	! AQ349923 RPCI-11-383F15.TV RPCI
gb_gss4:AQ307015	+	100.00	152.90	10.93	1003	! AQ307015 HS-2046.B1 H02-T7 CIT
gb_est44:AW603280	+	99.00	157.40	6.14	566	! AW603280 RCI-CN0012-170100-011
gb_gss16:B21398	-	99.00	133.10	10.66	842	! B21398 T8D17-Sp6 TAMU Arabidops
gb_est5:AW2194	-	96.50	145.90	26.83	1106	! AW2194 mc69a10.r1 Soares mous
gb_est44:AW609762	+	96.00	134.50	8.90	462	! AW609762 QV0-ST0247-200100-092
gb_est46:AW10271	-	96.00	153.95	9.55	486	! AW10271 eze08ne.r1 Neurospora
gb_gss11:AO751079	+	95.50	142.19	43.20	1332	! AO751079 HS-5576.B1 G04-SP6 RF
gb_est42:AA407899	+	95.00	134.45	8.96	397	! AA407899 UI-HF-BK0-ask-f-09-0-0
gb_est15:AI027770	+	94.50	145.49	28.46	839	! AI027770 ov99e09.x1 Soares_test
gb_gss12:AO843826	+	94.00	150.97	14.01	468	! AO843826 LMAJFV1_lm02d09.y3 Lei
gb_est42:AW402453	+	93.50	149.16	17.65	511	! AW402453 UI-HF-BK0-aan-f-12-0-0

gb\_gss4:AO324334 + 93.50 144.98 30.19 752 ! AO324334 mgxb0018E19r CUGI R  
gb\_est24:AI746782 + 93.00 151.96 12.33 365 ! AI746782 ul07d03.y1 Sugano m  
gb\_est44:AW604875 + 93.00 151.16 13.67 393 ! AW604875 QV1-CT0366-280100-0  
gb\_est46:AW726179 + 92.50 141.07 49.82 922 ! AW726179 GA\_Ea002006 Gossy

seq\_name: gb\_est42:AW402953

seq\_documentation\_block:  
LOCUS AW402953 312 bp mRNA EST 16-FEB-2000  
DEFINITION UI-HF-BK0-abd-c-06-0-UI.r1 NIH\_MGC\_36 Homo sapiens cDNA clone  
IMAGE:3055859 5', mRNA sequence.  
ACCESSION AW402953  
VERSION AW402953.1 GI:6921743  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 312)  
AUTHORS NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT On May 18, 1998 this sequence version replaced gi:3138550.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Eco RI site shown at the beginning of the sequence.  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
cDNA Library Preparation: M.B. Soares Lab  
cDNA Library Arrayed by: M.B. Soares Lab  
DNA Sequencing by: M.B. Soares Lab  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/dbrrp/image/image.html](http://www.bio.llnl.gov/dbrrp/image/image.html)  
Seq primer: M13 Forward.

FEATURES

source  
1..312  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3055859"  
/clone\_lib="NIH\_MGC\_36"  
/tissue\_type="lymph"  
/cell\_type="germinal center B cells"  
/lab\_host="DH10B (LT1)"  
/note="Vector: pT73-Pac; Site\_1: NotI; Site\_2: Eco RI;  
Constructed from size fractionated cytoplasmic mRNA  
(0.5-1.5kb). Directionally cloned. Cells provided by Louis  
M. Staudt, Ph.D. Library preparation by Maria de Fatima  
Bonaldio, Ph.D. and M. Bento Soares, Ph.D."

BASE COUNT 97 a 74 c 87 g 54 t

ORIGIN

alignment\_scores:

Quality: 544.00 Length: 101  
Ratio: 5.386 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-135-238B-2\_COPY\_18\_272 x AW402953

Align seg 1/1 to: AW402953 from: 1 to: 312

24 GluMethHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyth 40

|||||

10 GAATGATCATGTAGGATATATCTGCGGGGAGATGCTGGATCTGGAAC 59

|||||

40 rCysGlyThrValValSerThrThrAsnPhelLeuAlaGluTyrLysg 57

|||||

60 ATGTGTTACCGTGTGTATCCACCACCACTTCATCAAGGAGAATAACAAG 109

|||||

57 lyArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGlu 73

```
|||||
110 GCGAGTTACTCTGAAACATATACCCACGAAGATCTGTCTAGTGAG 159
74 ValThrGlnLeuThrGluSerAspSerGlyValThrAlaCysGlyAlaG1 90
|||||
160 GTAACACAGCTGACAGAAAGTGACAGCGGAGTCTATGCTGCGGAGCGG 209
90 yMetAsnThrAspArgGlyThrGlnLysValThrLeuAsnValHis 107
|||||
210 CATGAACACAGACCGGGAAAGACCCAGAAAGTCAACCTGAATGTCAC 259
107 erGluThrGluProSerTrpGluGluGlnProMetProGluThrProLys 123
|||||
260 GTGAATACGAGCCATCATGGGAGAGAGAGAGCAATGCTGAGAGACTC 309
124 Trp 124
|||
310 TGG 312
```

```
seq_name: gb_est45:AW655210
seq_documentation_block:
LOCUS AW655210 518 bp mRNA EST 05-APR-2000
DEFINITION 105863 MARC 180V Bos taurus cDNA 5', mRNA sequence.
ACCESSION AW655210
VERSION AW655210.1 GI:7421036
KEYWORDS EST.
SOURCE Bos taurus.
ORGANISM Bos taurus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 518)
Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A. and
Keele,J.W.
Design and use of four pooled tissue normalized cDNA libraries for
EST discovery in cattle
Unpublished (2000)
On Mar 8, 1999 this sequence version replaced gi:4388119.
Contact: Smith TPL
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904 e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCACGTCACGACG
Plate: 86 row: K column: 10
Seq primer: ATTAGTGACACTATAG.
```

```
FEATURES
source
1..518
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 180V"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site.1: XbaI; Site.2: XhoI;
Library made from pooled tissue from lymph node, ovary,
fat, hypothalamus, and pituitary."
BASE COUNT 124 a 133 c 157 g 104 t
ORIGIN
```

```
alignment_scores:
Quality: 451.00 Length: 119
Ratio: 4.295 Gaps: 0
Percent Similarity: 88.235 Percent Identity: 67.227
```

```
alignment_block:
US-09-135-238b-2_copy_18_272 x AW655210 ...
Align seg 1/1 to: AW655210 from: 1 to: 518
1 ArgIleLeuProGluValLysValGluGlyGluLeuGlyGlySerValTh 17
|||||
161 AAGGTCCTTCCAGAGTAAGATGAGGAGGAATGCTGGGAGGATCCATTAC 210
17 rIleLysCysProLeuProGluMetHisValArgIleTyrlLeuCysArg 34
|||||
211 CATCGAGTGCCTTCTGTAACGCGATGTGAGGATATATCTGTGCGCGA 260
34 LuMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsnPh 50
|||||
261 CCATGTACGAGTCTGGAAGATGCACACCGTGTGTCCAGCAACAATAC 310
51 lIleLysAlaGluTyrlLysGlyArgValThrLeuLysGlnTyrlProarg 57
|||||
311 CTCAGAGAGGAATTCAGACCCGAGTCACGCTGGAGCAGTGTCCGACAG 360
67 sAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerGly 84
|||||
361 GAATCTGTCTTGTGGTGATGCACAGAGCTGACCAAGAATGACAGCGGA 410
84 alTyrlAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLys 100
|||||
411 TCTATGCTGTGGGTGGCGGGAACACAGACCGAGCAAGCCAGCAGCAG 460
101 ValThrLeuAsnValHisSerGluTyrlGluProSerTrpGluGluGln 117
|||||
461 ATCACCCTGACTCTTCACAGCGTTTACGAGCCATCCTGGGAAGAAGACC 510
117 oMetPro 119
|||||
511 GATGCGCT 517
```

```
seq_name: gb_est42:AW444691
seq_documentation_block:
LOCUS AW444691 482 bp mRNA EST 17-FEB-2000
DEFINITION UI-H-B13-ajw-g-12-0-UI.sl NCI_CGAP_Sub5 Homo sapiens cDNA clone
IMAGE:2733502 3', mRNA sequence.
ACCESSION AW444691
VERSION AW444691.1 GI:6986453
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 482)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:
NCI-CGAP clone distribution information can be found through the
I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html The following repetitive
elements were found in this cDNA sequence: 1-30,
>POLY_A#Simple_repeat
Seq primer: M13 Forward
POLY_A=Yes.
Location/Qualifiers
1..482
/organism="Homo sapiens"
/db_xref="taxon:9606"
FEATURES
source
```

```
/clone_lib="NCI_CGAP_Sub5"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; NCI_CGAP_Sub5
is a subtracted library derived from NCI_CGAP_Sub4. The
NCI_CGAP_Sub5 library had 3 million recombinants. A
single-stranded DNA preparation of NCI_CGAP_Sub4 was used
as a tracer in a subtractive hybridization with a driver
comprising: the IMAGE pool (NCI_CGAP_Kid3 pool 1 LLAM
3334-3337, 3682-3683, 3798-3803 (IMAGE Clones)
1322376-1323911, 1456008-1456775,1500552-1502855);
NCI_CGAP_Kid5 pool 1 LLAM 3338-3342,3722-3725, 3776-3778
(IMAGE Clones) 1323912-1325831, 1471368-1472903,
1492104-1493255); NCI_CGAP_Lu5 pool 1 LLAM,3575-3582,
3851-3854 (IMAGE Clones)
1414920-1417991,1520904-1522439); NCI_CGAP_GC4 pool 1 LLAM
3164-3167, 3716-3720, 3733-3735 (IMAGE Clones)
1257096-1258631,1469064-1470983, 1475592-1476743);
NCI_CGAP_Pr22 pool 1 LLAM 2457-2459, 2758-2759, 3062-3068
(IMAGE Clones) 985608-986759,1101192-1101959,
1217928-1220615); NCI_CGAP_Co10 pool 1 LLAM 2644-2653,
2871-2872 (IMAGE Clones)
1057416-1061255,1144584-1145351). (10% of the driver
population), plus a pool of 3,840 arrayed clones from
NCI_CGAP_Sub1 (IMAGE Clones) 2708616-2710535) and
NCI_CGAP_Sub2 (IMAGE Clones) 2710536-2712455) (10% of
the driver population); plus a pool of 11,136 clones from
NCI_CGAP_Sub3 (IMAGE Clones) 2712456-2723591) (10% of the
driver population), plus a pool of 5,472 clones from
NCI_CGAP_Sub4 (IMAGE Clones) 2723592-2728969) (70% of the
driver population). Subtractions were performed as
previously described [Bonaldo, Lennon & Soares (1996):
Normalization and Subtraction: Two Approaches To
Facilitate Gene Discovery. Genome Research 6, 791-806. ;
TAG_LIB=NCI_CGAP_Kid5; TAG_TISSUE=kidney; TAG_SEQ=ATTC"
BASE COUNT 103 a 107 c 111 g 161 t
ORIGIN
```

```
alignment_scores:
  Quality: 444.50      Length: 117
  Ratio: 4.401        Gaps: 3
  Percent Similarity: 86.325  Percent Identity: 80.342
```

alignment\_block:

US-09-135-238b-2\_copy\_18\_272 x AW444691/rev ..

Align seg 1/1 to reverse of: AW444691 from: 1 to: 482

```
24  GluMetHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyTh 40
|||||
481  GNAATGATGTGAGGATATCTGTGCGGAGATGGCTGTGATCTGGAAC 432
|||||
40  rCysGlyThrValValSerThrAsnPhelileLysAlaGluTyrLysG 57
|||||
431  ATGTGGTACCGTGGTATCCACCACCACTTCATCAAGCACAATACRAGG 382
|||||
57  lYArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGlu 73
|||||
381  GCCGAGTTACTCTGAAGCAATACCCACCAAGAATCTGTTCTAGTGAGG 332
|||||
74  ValThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaG1 90
|||||
331  GTAACACAGCTGACAGAAGTGACAGCGGAGTCTATGCTCGGAGCGGG 282
|||||
90  yMetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnValHis 107
|||||
281  CATGAACACAGACCGGGGAAGACCCACAAGTCACTGATGTCCACA 232
|||||
107  er...GluTyrGluProSerTrpGluGlnProMetProGluThrPro 122
|| :|||
231  GTGGTAGGTCTCGGCTGATTGGAGGCTCAG.....ACTACCCAG 191
```

```
123  LysTrpPheHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSerSe 139
||| ||||| ||| :||| :||| :||| :||| :||| :||| :|||
190  AAAATGTTCACTTGGGAAGATTG...GAACAGCCATCACTGAACAG 144
139 r 139
143 T 143
```

seq\_name: gb\_est8:AA521993

```
seq_documentation_block:
LOCUS AA521993 627 bp mRNA EST 17-JUL-1997
DEFINITION vH61e12.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone
IMAGE:891502 5', mRNA sequence.
```

```
ACCESSION AA521993
VERSION AA521993.1 GI:2262738
SOURCE house mouse
ORGANISM Mus musculus
```

```
REFERENCE
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,K. and Waterston,K.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Jan 14, 1998 this sequence version replaced gi:1797443.
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MG1:519462
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 494.
```

FEATURES
source

```
1..627
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:891502"
/clone_lib="Soares_mammary_gland_NbMMG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
```

```
/note="Organ: mammary gland; Vector: pT7T3D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dr) primer [5',
TGTTCACCAATCTGAAGTGGAGCGCGCGCAATGGTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT7T3 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."
```

```
BASE COUNT 165 a 192 c 156 g 114 t
ORIGIN
```

```
alignment_scores:
  Quality: 425.00      Length: 184
  Ratio: 3.148        Gaps: 5
  Percent Similarity: 73.370  Percent Identity: 51.087
```





```

seq_documentation_block:
LOCUS      AW655218          494 bp      mRNA          EST          05-APR-2000
DEFINITION 105873 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION  AW655218
VERSION    AW655218.1  GI:7421044
KEYWORDS   EST.
SOURCE     Bos taurus.
ORGANISM   Bos taurus.

REFERENCE  Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
AUTHORS    Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A. and
           Keele,J.W.

TITLE      Design and use of four pooled tissue normalized cDNA libraries for
JOURNAL    EST discovery in cattle
COMMENT    EST discovered (2000)
           Unpublished (2000)
           On Mar 8, 1999 this sequence version replaced gi:4388128.
           Contact: Smith TPL
           USDA, ARS, US Meat Animal Research Center
           PO Box 166, Clay Center, NE 68933-0166, USA
           Tel: 402 762 4366
           Fax: 402 762 4390
           Email: smith@email.marc.usda.gov
           Single pass sequencing. Bases called and alt trimmed with phred
           v0.980904.e. Vector identified by cross_match with the -minscore 18
           and -minmatch 12 options.
           PCR primers
           FORWARD: AGGAACACGCTATGACCAT
           BACKWARD: GTTTCGCCAGTCACGACG
           Plate: 86 row: M column: 8
           Seq primer: ATTAGGTGCACACTATAG.
           Location/Qualifiers
             1..494
               /organism="Bos taurus"
               /db_xref="taxon:9913"
               /clone_lib="MARC 1BOV"
               /tissue_type="pooled"
               /lab_host="DH10B"
               /note="vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
               Library made from pooled tissue from lymph node, ovary,
               fat, hypothalamus, and pituitary."
BASE COUNT      118 a 125 c 150 g 101 t
ORIGIN

FEATURES             Location/Qualifiers
     source           1..494
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 1BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from lymph node, ovary,
                     fat, hypothalamus, and pituitary."
     BASE COUNT      118 a 125 c 150 g 101 t
     ORIGIN

alignment_scores:
  Quality: 384.00      Length: 111
  Ratio: 4.042         Gaps: 0
  Percent Similarity: 85.586 Percent Identity: 63.063

alignment_block:
us-09-135-238B-2_COPY_18_272 x AW655218  ..

Align seg 1/1 to: AW655218 from: 1 to: 494

1 ArgfileLeuProGluValLysGluLeuGlySerValTh 17
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
162 AGGTCTCTCCAGAGTAAGATGGAAGGAAATGCTGGGAGGATCCATTAC 211
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|

17 rleLysCysProLeuProGluMethHisValArgIleTyrLeuCysArg 34
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
212 CATCGAGTCCCACTCTCTGAAACGATCTGAGGATATATCTGTGCCGGA 261
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|

34 luMethAlaGlySerGlyThrCysGlyThrValValSerThrThrAsnPhe 50
...|.....|.....|.....|.....|.....|.....|.....|.....|.....|
262 CCATTGACGAGTCTGGAGATGACACCACCGTGTCTCCAGCAACAATAC 311
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|

51 IleLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProArgLy 67
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
312 GTNAGAGAGATTCAAGCACCGGACGCTGGAGCAGCTGTCGGACAG 361
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|

```

```

67 sasLeuPheLeuValGluValThrGlnLeuLeuThrGluSerAspSerGlyV 84
:|||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
362 GAATCTGTCTCTGTGGTGACAGAGCTGACCAAGATGACAGCGGA 411
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|

84 alTyAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLys 100
:|||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
412 TCTATGCTGTGGGTGGCGGGAACACAGACCGAGCGGACAGACCCATCAG 461
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|

101 ValThrLeuAsnValHisSerGluTyrGluPro 111
:|||||.....|.....|.....|.....|.....|.....|.....|.....|.....|
462 ATCACCTGACTGTTACAGCGTTTACTAGCCA 494
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|

seq_name: gb_est25:AI760373

seq_documentation_block:
LOCUS      AI760373          711 bp      mRNA          EST          20-DEC-1999
DEFINITION wh87d11.x1 NCI_CGAP_CLL1 Homo sapiens cDNA clone IMAGE:2387733 3'
           similar to TR:O60667 O60667 ANTI-FAS-INDUCED APOPTOSIS. ; mRNA
           sequence.
ACCESSION  AI760373
VERSION    AI760373.1  GI:5176040
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 711)
AUTHORS    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
           National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
           Tumor Gene Index
           Unpublished (1997)
JOURNAL
COMMENT     On Oct 30, 1998 this sequence version replaced gi:3812483.
           Contact: Robert Strausberg, Ph.D.
           Tel: (301) 496-1550
           Email: Robert_Strausberg@nih.gov
           Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever,
           M.D., Louis M. Staudt, M.D., Ph.D.
           CDNA Library Preparation: M. Bento Soares, Ph.D.
           CDNA Library Arrayed by: Greg Lennon, Ph.D.
           DNA Sequencing by: Washington University Genome Sequencing Center
           Clone distribution: NCI-CGAP clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           www-bio.llnl.gov/bbrp/image/image.html
           Insert Length: 1449 Std Error: 0.00
           Seq primer: -400P from Gibco
           High quality sequence stop: 472.
           Location/Qualifiers
             1..711
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone_image="IMAGE:2387733"
               /clone_lib="NCI_CGAP_CLL1"
               /tissue_type="B-cell, chronic lymphocytic leukemia"
               /lab_host="DH10B"
               /note="vector: pT7T3D-Pac (Pharmacia) with a modified
               polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
               was primed with a Not I - oligo(dT) primer [5',
               TGTTACCAATCTGAAGTGGAGCGCGCATGCTGTTTTTTTTTTTTTTTTTTT
               T 3']; double-stranded cDNA was ligated to Eco RI
               adaptors (Pharmacia), digested with Not I and cloned into
               the Not I and Eco RI sites of the modified pT7T3 vector.
               Library is normalized, and was constructed by Bento
               Soares and M.Fatima Bonaldo."
BASE COUNT      127 a 195 c 239 g 148 t
ORIGIN

alignment_scores:
  Quality: 328.00      Length: 84
  Ratio: 4.316         Gaps: 1
  Percent Similarity: 90.476 Percent Identity: 83.333

```

## alignment\_block:

US-09-135-238B-2\_COPY\_18\_272 x A1760373/rev ..

Align seg 1/1 to reverse of: A1760373 from: 1 to: 711

```

173 valSerArgAlaSerSerValAlaGlyAspLysProArgThrPheLeuPr 189
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
711 GTGTCAAGACATTCTCAAGTAGCAGGTGACAAAGCNCGACCTCTTTGCC 662

189 oSerThrAlaSerLysIleSerAlaLeuGluGlyLeuLeuLysProG 206
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
561 ATCCACTACAGCTCAAAATCTCAGCTCTGGAGGGGCTGCTCAAGCCCC 612
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
206 lnThrPro_SerTyrAsnHisHisThrArgLeuHisArgGlnAlaLe 222
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
611 AGACGCCAGGGCTACACACACCACCAGCGCTGCACAGCGAGAGCACT 562

222 uAspTyrGlySerGlnSerGlyArgGluGlyGlnGlyPheHisIleLeu 239
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
561 GGACTAAGCTCACAGTCTGGAGGAGGACCAAGGATTCACATCCTGA 512

239 leProThrIleLeuGlyLeuPheLeuAlaLeuLeuGlyLeuValVal 255
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
511 TCCCGACCATCTGGGGCTTTTCTGCTGGCACTTCTGGGGTGGTGGTG 462

```

seq\_name: gb\_est12:AA830944

## seq\_documentation\_block:

```

LOCUS      AA830944      485 bp      mRNA      EST      07-APR-1998
DEFINITION OC58B10.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1353883 3',
            mRNA sequence.
ACCESSION  AA830944
VERSION    AA830944.1 GI:2904043
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 485)
AUTHORS    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL    Unpublished (1997)
COMMENT    Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert.Strausberg@nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            cDNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 983 Std Error: 0.00
            Seq primer: -40ml3 fwd. Et from Amersham
            High quality sequence stop: 473.
            Location/Qualifiers
                1..485
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:1353883"
                    /clone_lib="NCI_CGAP_GCB1"
                    /tissue_type="germinal center B cell"
                    /lab_host="DH10B"
                    /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                    polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA
                    was prepared from human tonsillar cells enriched for
                    germinal center B cells by flow sorting (CD20+, IgD-),
                    provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                    (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
                    primed with a Not I - oligo(dT) primer

```

## FEATURES

```

source
1..179
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:824022"

FEATURES
             source
1..179
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:824022"

[5'-TGTACCAATCTGAAGTGGAGCGCGCTCATTTTTTTTTTTTTTTT-3'
]. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      109 a      84 c      137 g      155 t
ORIGIN

alignment_scores:
    Quality:      305.00      Length:      59
    Ratio:        5.169      Gaps:      0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-135-238B-2_COPY_18_272 x AA830944/rev ..
Align seg 1/1 to reverse of: AA830944 from: 1 to: 485

162 SerProThrThrGlnIleThrHisArgProArgValSerArgAlaSerSe 178
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
485 TCCCCCACCAACCAATCAACCCAGCGCTCGAGTGTCCAGAGCATCTTC 436
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
178 rValAlaGlyAspLysProArgThrPheLeuProSerThrThrAlaSerL 195
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
435 AGTAGCAGGTGACAAAGCCCGAAGCTTCTGTCATCCACTACAGCCTCAA 386
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
195 ysIleSerAlaLeuGluGlyLeuLysProGlnThrProSerTyrAsn 211
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
385 AAATCTCAGTCTGGAGGGGCTGCTCAAGCCCGACGCGCCAGGTACAC 336
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
212 HisHisThrArgLeuHisArgGlnArg 220
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
335 CACCACACAGGCTGCACAGGCAGAGG 309

seq_name: gb_est8:AA491123
seq_documentation_block:
LOCUS      AA491123      179 bp      mRNA      EST      18-AUG-1997
DEFINITION aa4604.r1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:824022 5',
            mRNA sequence.
ACCESSION  AA491123
VERSION    AA491123.1 GI:2220296
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 179)
AUTHORS    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL    Unpublished (1997)
COMMENT    Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert.Strausberg@nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -28ml3 rev1 Et from Amersham
            High quality sequence stop: 116.
            Location/Qualifiers
                1..179
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:824022"

```

```
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD+),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTCACATCTGAAGTGGAGCGCCCTCATTTTTTTTTTTTTTTT-3'
]. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
```

BASE COUNT 46 a 69 c 34 g 30 t

ORIGIN

alignment\_scores:

Quality:	299.00	Length:	59
Ratio:	5.068	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000

alignment\_block:

US-09-135-238B-2\_COPY\_18\_272 x AA491123

Align seg 1/1 to: AA491123 from: 1 to: 179

146 ValThrThrProAlaGlnArgGlyLysValProValHisHisSerSe 162  
|||||

3 GTTACCACACCAGCTCAAGGGGCAAGTCCCTCCAGTTCACCATCTCTC 52  
|||||

162 rProThrThrGlnIleThrHisArgProArgValSerArgAlaSerV 179  
|||||

53 CCCCACACCCCAAAATCACCCACCGACCTCGAGTGTCCAGACATCTTCAG 102  
|||||

179 aAlaGlyAspLysProArgThrPheLeuProSerThrThrAlaSerLys 195  
|||||

103 TAGCAGGTGCACAGCCCGCAACCTTCCTGCCATCCACTACAGCCTCAAAA 152  
|||||

196 IleSerAlaLeuGluGlyLeuLeuLys 204  
|||||

153 ATCTCAGCTCTGGAGGGGTGCTCAAG 179  
|||||

seq\_name: gb\_est40:AW291041

seq\_documentation\_block:

LOCUS AW291041 561 bp mRNA EST 16-JAN-2000

DEFINITION UT-H-B12-ag1-h-06-0-UI.sl NCI\_CGAP\_Sub4 Homo sapiens cDNA clone  
IMAGE:2724707 3', mRNA sequence.

ACCESSION AW291041

VERSION AW291041.1 GI:6697677

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 561)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On May 18, 1998 this sequence version replaced gi:3137211.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
The sequence contained an oligo-dT track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. cDNA Library Preparation: M.B. Soares lab clone distribution:  
NCI-CGAP clone distribution information can be found through the

I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/dbbrp/image/image.html  
Seq primer: M13 Forward  
POLYA=Yes.

FEATURES

source

1..561

Location/Qualifiers

organism="Homo sapiens"  
db\_xref="taxon:9606"  
clone="IMAGE:2724707"  
clone\_lib="NCI\_CGAP\_Sub4"  
lab\_host="DH10B (Life Technologies)"  
note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; The  
NCI\_CGAP\_Sub4 library is a subtracted library derived from  
the NCI\_CGAP\_Sub2 library which is a subtracted library  
derived from the NCI\_CGAP\_Sub1 library, which is a  
subtracted library derived from B1. B1 constitutes a  
mixture of 21 normalized or subtracted NCI\_CGAP  
libraries: NCI\_CGAP\_Co4, NCI\_CGAP\_Pr22, NCI\_CGAP\_Pr28,  
NCI\_CGAP\_Co10, NCI\_CGAP\_Co16, NCI\_CGAP\_Kid5,  
NCI\_CGAP\_Kid12, NCI\_CGAP\_Kid3, NCI\_CGAP\_Kid11,  
NCI\_CGAP\_Lym2, NCI\_CGAP\_Br2, NCI\_CGAP\_Co8, NCI\_CGAP\_CLL1,  
NCI\_CGAP\_Le12, NCI\_CGAP\_Brn23, NCI\_CGAP\_Lu5,  
NCI\_CGAP\_Lu24, NCI\_CGAP\_Lu19, NCI\_CGAP\_GC4, NCI\_CGAP\_GC6,  
NCI\_CGAP\_Brn25. These 21 libraries were pooled and a  
single-stranded DNA preparation of the resulting mixture  
was used as a tracer in a subtractive hybridization with  
a driver whose composition is detailed below:  
NCI\_CGAP\_Kid3 pool 1 : LLAM 3334-3337, 3682-3683,  
3798-3803 (IMAGE CloneIDs 1322376-1323911,  
1456008-1456775, 1500552-1502855) NCI\_CGAP\_Kid5 pool 1 :  
LLAM 3338-3342, 3722-3725, 3776-3778 (IMAGE CloneIDs  
132912-1329831, 1471368-1472903, 1492104-1493255)  
NCI\_CGAP\_Lu5 pool 1 : LLAM 3575-3582, 3851-3854 (IMAGE  
CloneIDs 1414920-1417991, 1520904-1522439) NCI\_CGAP\_GC4  
pool 1 : LLAM 3164-3167, 3716-3720, 3733-3735 (IMAGE  
CloneIDs 1257096-1258631, 1468064-1470983,  
1475592-1476743) NCI\_CGAP\_Pr22 pool 1 : LLAM 2457-2459,  
2758-2759, 3062-3068 (IMAGE CloneIDs 985608-986759,  
1101192-1101959, 1217928-1220615) NCI\_CGAP\_Co10 pool 1 :  
LLAM 2644-2653, 2871-2872 (IMAGE CloneIDs 1057416-1061255,  
1144584-1145351) Subtraction was performed as previously  
described (Bonaldo, Lennon & Soares (1996): Normalization  
and Subtraction: Two Approaches to Facilitate Gene  
Discovery. Genome Research 6, 791-806.);  
TAG\_LIB=NCI\_CGAP\_Co10; TAG\_TISSUE=colon; TAG\_SEQ=AAACG"

BASE COUNT 121 a 125 c 120 g 194 t 1 others

ORIGIN

alignment\_scores:

Quality:	292.50	Length:	91
Ratio:	3.953	Gaps:	3
Percent Similarity:	81.319	Percent Identity:	72.527

alignment\_block:

US-09-135-238B-2\_COPY\_18\_272 x AW291041/rev

Align seg 1/1 to reverse of: AW291041 from: 1 to: 561

50 PheIleLysAlaGluThrLysGlyArgValThrLeuLysGluThrProAr 66  
|||||

561 TTCATCAAGCAGATACAGCCCGAGTCTCTGAAGCAATACCCACG 512  
|||||

66 gLysAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerG 83  
|||||

511 CAAGAATCTGTTCTTAGTGGAGTACACAGCTGACAGAAAGTGACACG 462  
|||||

83 lyValThrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGln 99  
|||||

461 GAGTCTATGCTCGCGAGCGGCATGAACACAGACCGGNGAAGACCCAG 412  
|||||

100 LysValThrLeuAsnValHisSer...GluTyrGluProSerTrpGluCl 115

```

|||||
411 AAGTCACCTGAATGTCACAGTGGTAGGTCCTCCGCTGATTGGAGGC 362
|||||
115 uGlnProMetProGluThrProLysTrpPheHisLeuProTyrLeuPheG 132
|||||
361 TCAG.....ACTACCCAGAAAATGTTTCACCTGGGAAGATTG...G 324
|||||
132 lnMetProAlaTyrAlaSerSer 139
|||||
323 AACAGCCATCTAACTGTAACAGT 301
|||||

```

seq\_name: gb\_est4:AA289667

seq\_documentation\_block: 417 bp mRNA EST 14-APR-1997  
 LOCUS AA289667 v20c07.r1 Soares mouse 3NDMS Mus musculus cDNA clone IMAGE:749484  
 DEFINITION 5', mRNA sequence.

ACCESSION AA289667  
 VERSION AA289667.1 GI:1937343

KEYWORDS EST.  
 SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE 1 (bases 1 to 417)  
 AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, F., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

TITLE The WashU-HMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of MedicineP

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.

MG1:458468

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 405.

FEATURES

source

1. .417

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="IMAGE:749484"

/clone\_lib="Soares mouse 3NDMS"

/sex="male"

/tissue\_type="Spleen"

/dev\_stage="4 weeks"

/lab\_host="DH10B"

/notes="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker; Site.1: Not 1; Site.2: Eco RI; 1st strand cDNA  
 was primed with a Not I - oligo(dT) primer [5,

TGTTACCAATCTGAAGTGGAGCGCGCTGTTTTTTTTTTTTTTTTTTT

3.); double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT73 vector. RNA

provided by Dr. Bertrand Jordan. Library went through

three rounds of normalization, and was constructed by

Bento Soares and M.Fatima Bonaldo."

BASE COUNT 109 a 140 c 91 g 77 t

ORIGIN

alignment\_scores:

Quality: 281.00 Length: 118

Ratio: 3.122 Gaps: 2

Percent Similarity: 76.271 Percent Identity: 50.847

alignment\_block:

US-09-135-238B-2\_COPY\_18\_272 x AA289667 ..

Align seg 1/1 to: AA289667 from: 1 to: 417

```

145 ArgValThrProAlaGlnArgGlyLysValProProValHisHis 161
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4 AAAGTTACACGCCCAACTCCAAAGACTGAGGCCCTCGGTTCCACG 53

```

```

161 rSerProThrThrGlnIleThrHisArgProArgValSerArgAlaSer 178
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
54 CTCGAGCATCATCTTCAGTAAACCAATCCAGAGTTTACAGAGCATTTT 103

```

```

178 erValAlaGlyAspLysProArgThrPheLeuProSerThrThrAlaSer 194
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
104 CTGTGTCAGTCAGTACCAGTCCCAAGCTCTCCAGCAACACACAGCTCA 153

```

```

195 LysIleSerAlaLeuGluGlyLeuLysProGlnThrProSerTyrAs 211
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
154 AAGACTTCCACTCAGCAAGCA...ATCAGGCCCTAGAGCCAGCTACAG 200

```

```

211 nHisHisThrArgLeuHisArgGlnArgAlaLeuAspTyrGlySerGln 228
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
201 CCACCAACACCAAGCTTCATGAGCAAGAGACGCCACCATGGGCCACACT 250

```

```

228 erGlyArgGluGlyGlnGly.....PheHisIle 237
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
251 ATGGGAGAGAAAGACCGGGCTTCACATCCCATCCAGAAATTCACATC 300

```

```

238 LeuIleProThrIleLeuGlyLeuPheLeuAlaLeuLeuGlyLeuVa 254
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
301 CTGATCCGACCTCTCGGGCTTCTCTCTGCTGGTCTTTTGGGACTGGT 350

```

254 lVal 255

|||||

351 GGTA 354

seq\_name: gb\_est3:AA174968

seq\_documentation\_block:

LOCUS AA174968 460 bp mRNA EST 16-FEB-1997  
 DEFINITION ms78h03.r1 Soares mouse 3NDMS Mus musculus cDNA clone IMAGE:617717  
 5', mRNA sequence.

ACCESSION AA174968

VERSION AA174968.1 GI:1756098

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE 1 (bases 1 to 460)

AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, F., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

TITLE The WashU-HMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of MedicineP

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MG1:378541

Seq primer: -28M13 rev2 from Amersham

High quality sequence stop: 459.

FEATURES

source

Location/Qualifiers

1. .460

```
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db_xref="taxon:10090"  
/clone_image:61717"  
/clone_lib="Soares mouse 3Nbms"  
/seq="malon"
```

/notes/Vector: pT73D-Pac (Pharmacia) with a modified polylinker: Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I oligo(dT) primer [5', TGTTACATCTGATGAGCGACGCCGCGCTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT733 vector. RNA provided by Dr. Bertrand Jordan. Library went through three rounds of normalization, and was constructed by Bento Soares and M.Fátima Bonaldo."

BASE COUNT	108 a	158 c	117 g	77 t
ORIGIN				
alignment_scores:				
Quality:	218.00		Length:	92
Ratio:	3.159		Gaps:	2
Percent Similarity:	75.000		Percent Identity:	53.261

alignment\_block:  
US-09-135-238B-2\_COPY\_18\_272 x AA174968 ..

Align seg 1/1 to: AA174968 from: 1 to: 460

171 ProArgValSerArgAlaSerSerValAlaGlyAspLysProArgThrPh 187

7 CCCAGAGTTACAGAGCATTTCTGTGTGCTACCAAGTCCCCAGCGCT 56

187 eLeuProSerThrThrAlaSerLysIleSerAlaLeuGluGlyLeuLeuL 204

57 CCTGCCAGCAACCAAGCCTCAAGACTTCCACTCAGCAAGCA .ATCA 103

204 ysProGlnThrProSerTyrAsnHisHisThrArgLeuHisArgGlnArg 220

104 GGCCCCTAGAGGCCAGCTACAGCCACCACCCAGACTTCATGAGCAAGG 153

221 AlaLeuAspTyrGlySerGlnSerGlyArgGluGlyGlnGly..... 234

154 ACACGCCACCATGGCCACACTATGGGAGAGAAGACCGGGCTTCACAT 203

235 .....PheHisIleLeuIleProThrIleLeuGlyLeuPheL 247

204 CCCCATCCCAGAATTTCACATCCTGATTCGACCTTCCTGGGCTTTCTCT 253

247 euLeuAlaLeuLeuGlyLeuValVal 255

254 TGCTGGTCTTTGGACTGGTGGTA 279

seq\_name: gb\_est8:AA509857

seq_documentation_block:	199 bp	mrna	EST	08-JUL-1997
LOCUS	AA509857			
DEFINITION	W51410.1 Soares_mammary_gland_NBMWG	Mus musculus	cdna clone	
	IMAGE:890490	5' similar to SW:PIGR_HUMAN P01833		
		POLYMERIC-IMMUNOGLOBULIN RECEPTOR ;	mrna sequence.	

Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

BASE COUNT	55 a	40 c	58 g	46 t
ORIGIN				
alignment_scores:				
Quality:	209.50			Length: 67
Ratio:	3.741			Gaps: 1
Percent Similarity:	83.582			Percent Identity: 62.687

alignment\_block:

US-09-135-238B-2\_COPY\_18\_272 x AA509857

Align seg 1/1 to: AA509857 from: 1 to: 199

27 ValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysGlyTh 43

2 GTAAGGATGTATCTGTGTCGGCAGATGGCCAAACCTGGGATATGCTCCAC 51

43 rValValSerThrThrAsnPhelLeLysAlaGluTyrLysGlyArgValT 60

52 TGTGGTGTCCAACACC...TTTGTCAAGAAAGGAATATGAAAGGCGAGTCA 98

60 hrLeuLysGlnTyrProArgIysAsnIleuPheIleuValGlnuValThrGln 76

99 CCTGAGGCCATGCTTGGTATACACGCTATTCCTATCTCCACATCCACAC 148

```
77 LengthRC[usor] = usor[usorLen];
```

[illegible]

5  
6  
7  
8

199 A 199  
seq\_name: gb\_est36:AW012833  
seq\_documentation\_block:  
LOCUS AW012833 327 bp mRNA EST 10-SEP-1999  
DEFINITION uc034111.v1 Sugano mouse liver mlia Mus musculus cDNA clone  
IMAGE:2582325 5' similar to gb:U06431 Mus musculus B6/CBAF1J  
polymERIC immunoglobulin receptor mRNA, (MOUSE);, mRNA sequence.  
ACCESSION AW012833  
VERSION AW012833.1 GI:5861611  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 327)  
Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,  
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,  
Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,  
Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,  
Waterston,R. and Wilson,R.  
The WashU-NCI Mouse EST Project 1999  
Unpublished (1999)  
On Jul 9, 1999 this sequence version replaced gi:5433821.  
Contact: Marra M/WashU-NCI Mouse EST Project 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:1021913  
Seq primer: custom primer used  
High quality sequence stop: 272.  
FEATURES  
source  
1..327  
/organism="Mus musculus"  
/strain="C57BL"  
/db\_xref="taxon:10090"  
/clone="IMAGE:2582325"  
/clone\_lib="Sugano mouse liver mlia"  
/sex="female"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Organ: liver; Vector: pME18S-FL3; Site:1: DraIII  
(CACCATGTG); Site\_2: DraIII (CACCATGTG); 1st strand cDNA  
was primed with an oligo(dT) primer  
[ATGTGGCTTTTCTTTTCTTTTCTTTT]; double-stranded cDNA was  
ligated to a DraIII adaptor [TGTGGCCTACTGG], digested  
and cloned into distinct DraIII sites of the pME18S-FL3  
vector (5' site CACCATGTG, 3' site CACCATGTG). XhoI should  
be used to isolate the cDNA insert. Size selection was  
performed to exclude fragments <1.5kb. Library  
constructed by Dr. Sumio Sugano (University of Tokyo  
Institute of Medical Science). Custom primers for  
sequencing: 5' end primer CTTCGCTCTTAAAGGTGG and 3' end  
primer CGACCTGCAGCTCGACACA."  
BASE COUNT 87 a 98 c 76 g 66 t  
ORIGIN  
alignment\_scores:  
Quality: 180.00 Length: 88  
Ratio: 2.769 Gaps: 3  
Percent Similarity: 73.864 Percent Identity: 40.909  
alignment\_block:  
US-09-135-238b-2\_COPY18\_272 x AW012833 ..  
Align seg 1/1 to: AW012833 from: 1 to: 327

13 GlyGlySerValThrIleLysCysProLeuProGlu.....Me 25  
||| ||||| ||||| ||| |||||  
60 GGCAGACTCTGTTCCATCAGCTGTCTACTACCCAGACACCTCTGTCAACCG 109  
||| ||||| ||||| ||| |||||  
25 thisValArgIleTyrLeuCysArgGluMetaIaGlySerGlyThrCysG 42  
||| ||||| ||| ||||| ||||| ||| ||||| |||||  
110 GCACACCCCGGAATACTGGTGGCCGACANGAGCC....AGCGCATGTGCA 156  
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
42 lyThrValValSerThrThrAsnPhelIleLysAlaGluTyrLysGlyArg 58  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
157 CAACGCTCATCTCTTCAAAATGGCTACCTCTCCAAAGGAGTATTTCAGCAGA 206  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
207 GCCAACCTCATCACTCTCCAGAGAAACACACATTGTGATTACATTGA 256  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 92  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
257 GCAGCTCACCAGGACGACACTGGGAGCTACAGGTGGCTGGGT...A 303  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
92 snThrAspArgGly 96  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
304 CCAGCAACCGAGGC 317  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
seq\_name: gb\_est15:AI021484  
seq\_documentation\_block:  
LOCUS AI021484 220 bp mRNA EST 16-JUN-1998  
DEFINITION ub08h03.r1 Soares\_mammary\_gland\_NbMMG Mus musculus cDNA clone  
IMAGE:1366421 5', mRNA sequence.  
ACCESSION AI021484  
VERSION AI021484.1 GI:3235820  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 220)  
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
Waterston,R.  
The WashU-HMI Mouse EST Project  
Unpublished (1996)  
On Jun 5, 1998 this sequence version replaced gi:3186958.  
Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:899641  
Seq primer: -28m13 rev2 ET from Amersham  
High quality sequence stop: 141.  
FEATURES  
Location/Qualifiers  
1..220  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:1366421"  
/clone\_lib="Soares\_mammary\_gland\_NbMMG"  
/sex="male"  
/cissue\_type="mammary gland"  
/dev\_stage="4 weeks"  
/lab\_host="DH10B"







OM of: US-09-135-238B-2\_COPY\_273\_390 to: GenEmbl:\* out\_format : ffs  
Date: Sep 12, 2000 6:56 AM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:  
-MODEL=frame-p2n.model -DEV=xlp  
-O=/cgn2\_1/USPTO\_spo1/US09135238/runat\_24082000\_091323\_17880/app\_query.fasta\_1.1282  
-DB=GenEmbl -OEMT=fastap -SUFFI=p2n.rge -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000  
-QAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blossum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ffs  
-NORM=ext -MINLEN=0 -MAXLEN=1000000  
-USER=US09135238@cgn1\_1\_5120 -NCPU=6 -ICPU=3 -LONGLOG -NO\_XLPXY  
-WAIT -THREADS=1

Search information block:  
Query: US-09-135-238B-2\_COPY\_273\_390  
Query length: 118  
Database: GenEmbl:\*  
Database sequences: 972840  
Database length: 892348106  
Search time (sec): 2784.300000

score_list:	Strd	Orig	Zscore	EScore	Len	Documentation
gb_pr3:AF057557	+	616.00	637.82	2.0e+27	1339	AF057557 Homo sapiens anti-Fas
gb_htg11:AC023534	+	325.00	308.02	4.6e+09	221365	AC023534 Homo sapiens clone
gb_htg20:AC010538	-	107.00	85.28	1.2e+04	249227	AC010538 Homo sapiens chromo
gb_ba1:SC135	-	105.00	95.10	3.4e+03	490909	AL031541 Streptomyces coelic
gb_htg14:AC025634	+	104.50	91.25	5.1e+03	68092	AC025634 Homo sapiens chromo
gb_pr4:AC020663	+	104.50	85.89	1.1e+04	154218	AC020663 Homo sapiens chromo
gb_htg22:AC027687	+	104.50	82.83	1.6e+04	254472	AC027687 Homo sapiens chromo
gb_ba1:SCP43A	-	104.00	95.03	3.4e+03	35437	AL096837 Streptomyces coelic
gb_ba1:SVPKS	-	103.50	106.38	790.84	5820	X16300 Streptomyces violaceoru
gb_pr1:AB019004	-	103.50	105.42	893.49	6728	AB019004 Homo sapiens DNA, 5'
gb_ba1:SVGRA	-	103.50	105.36	900.75	6793	X16144 S. violaceoruber DNA f
gb_ro:RTPRMTNP2	+	103.50	101.01	1.6e+03	13187	Z46939 R. norvegicus PRM1, PR
gb_ba1:SCB63	+	103.50	94.20	3.8e+03	37200	AL035540 Streptomyces coelic
gb_ba1:SVT011500	-	103.50	93.85	3.9e+03	32550	AJ011500 Streptomyces violac
gb_pr3:HS16389	-	103.50	85.35	1.2e+04	143291	AJ008733 Human DNA sequence
gb_pr1:HUMFGRB	-	103.00	114.54	277.63	1553	L78720 Human fibroblast growth
gb_htg8:AC020867	+	103.00	114.35	284.23	1597	L12060 Homo sapiens retinoid a
gb_ba2:AE001874	-	102.50	101.19	1.5e+03	10991	AF001874 Deinococcus radiodur
gb_pr2:HS74782	+	102.00	85.57	1.1e+04	109859	AJ021393 Human DNA sequence
gb_ba2:AF071081	+	101.50	106.89	740.36	3946	AF071081 Mycobacterium tubercu
gb_pat:AG3410	+	101.50	106.89	740.36	3946	A93410 Sequence 6 from Patent
gb_ba1:SCNUSG	-	101.50	102.91	1.2e+03	7235	X72787 S. griseus nusG, rplKJL
gb_pr4:AC004973	-	101.50	85.87	1.1e+04	97037	AC004973 Homo sapiens PAC clc
gb_pr3:HUMYWD703	-	101.50	83.70	1.4e+04	135038	L78810 Homo sapiens ADP/ATP
gb_ro:AF082568	+	101.00	103.67	1.1e+03	5968	AF082568 Mus musculus zinc fin
gb_htg5:AC016278	+	101.00	91.90	5.1e+03	35816	AC016278 Homo sapiens clone R
gb_ba1:SC5C7	+	101.00	90.87	5.8e+03	41906	AL031515 Streptomyces coelic
gb_ba1:SCW1	-	101.00	90.56	6.0e+03	43961	AL13422 Streptomyces coelic
gb_htg27:AC022536	+	101.00	81.89	1.8e+04	164587	AC022536 Homo sapiens clone
gb_htg18:AC025300	-	101.00	81.74	1.9e+04	168628	AC025300 Homo sapiens chromo
gb_pr4:AC004169	-	101.00	80.96	2.0e+04	198805	AC004169 Homo sapiens chromo
gb_htg20:AC024549	+	101.00	79.96	2.3e+04	220961	AC024549 Homo sapiens chromo
gb_pr4:AF152486	+	100.50	109.01	563.85	2445	AF152486 Homo sapiens protocad
gb_pr4:AF152316	+	100.50	103.99	1.1e+03	5260	AF152316 Homo sapiens protocad
gb_ba1:SC6D7	-	100.50	91.05	5.6e+03	37750	AL133213 Streptomyces coelic
gb_htg18:AC026593	-	100.50	89.41	7.0e+03	48459	AC026593 Homo sapiens chromo
gb_htg20:AC008468	+	100.50	83.47	1.5e+04	119783	AC008468 Homo sapiens chromo
gb_pr3:AC005609	-	100.50	81.65	1.9e+04	137970	AC005609 Homo sapiens chromo
gb_htg2:AL157817	-	100.50	80.65	2.1e+04	184106	AL157817 Homo sapiens chromo

gb\_pr1:HUMHRARC + 100.00 110.15 487.42 1903 ! M24857 Human retinoic acid  
gb\_ba1:STMSEC - 100.00 109.91 502.89 1975 ! D17464 S. griseus genes for  
gb\_pr1:AB018254 - 100.00 101.88 1.4e+03 6706 ! AB018254 Homo sapiens mRNA  
gb\_ba1:SC9A10 - 100.00 99.95 1.8e+03 9000 ! AL031260 Streptomyces coeli  
gb\_ba1:MTCX10D7 - 100.00 90.19 6.3e+03 39800 ! 279700 Mycobacterium tuber  
seq\_name: gb\_pr3:AF057557  
seq\_documentation\_block:  
LOCUS AF057557 1339 bp mRNA PRI 04-JUN-1998  
DEFINITION Homo sapiens anti-Fas-induced apoptosis (TOSO) mRNA, complete cds.  
ACCESSION AF057557  
VERSION AF057557.1 GI:3169292  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1339)  
AUTHORS Hitoshi,Y., Lorens,J., Kitada,S.I., Fisher,J., LaBarge,M.,  
Ring,H.Z., Francke,U., Reed,J.C., Kinoshita,S. and Nolan,G.P.  
TITLE Toso, a cell surface, specific regulator of Fas-induced apoptosis  
in T cells  
JOURNAL Immunity 8 (4), 461-471 (1998)  
MEDLINE 98246048  
REFERENCE 2 (bases 1 to 1339)  
AUTHORS LaBarge,M. and Hitoshi,Y.  
TITLE Direct Submission  
JOURNAL Submitted (03-APR-1998) Molecular Pharmacology, Stanford  
University, 300 Pasteur Dr., Palo Alto, CA 94304, USA  
FEATURES  
Location/Qualifiers  
1..1339  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1q31-q32"  
/cell\_line="Jurkat"  
/cell\_type="T cell"  
/note="TOSO.2 cDNA; cloned by PCR using primers designed  
from the original TOSO.1 cDNA sequence"  
1..1339  
/gene="TOSO"  
20..1192  
/gene="TOSO"  
/note="transmembrane protein; Ig superfamily member"  
/codon\_start=1  
/product="anti-Fas-induced apoptosis"  
/protein\_id="AAC18830.1"  
/db\_xref="GI:3169293"  
variation  
127  
/gene="TOSO"  
/note="occurs in TOSO.1 cDNA"  
/replace="a"  
1015  
/gene="TOSO"  
/note="occurs in TOSO.1 cDNA"  
/replace="g" 340 g 258 t  
BASE COUNT 309 a 432 c 340 g 258 t  
ORIGIN  
alignment\_scores:  
Quality: 616.00 Length: 118  
Ratio: 5.220 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:

US-09-135-238B-2 COPY 273 390 X AF057557

Align seq 1/1 to: AF057557 from: 1 to: 1339

1 LysArgAlaValGluArgGlyLysAlaLeuSerArgAlaArgIle 17  
836 AAAAGGGCGTTGAAAGGAGAAAGCCCTCCAGCGCGGCGCCGCGACT 885  
17 uAlaValArgMetArgAlaLeuGluSerSerGlnArgProArgGlySerP 34  
886 GGCCTGAGGATGCGCGCTCGAGACTCCAGAGGCCCGCGGGTGCG 935  
34 roArgProArgSerGlnAsnAsnIleTyrSerAlaCysProArgArgAla 50  
936 CGGAGCGCGCTCCAAAAACAACATCTACAGCGCTCGCGGCGCGCT 985  
51 ArgGlyAlaAspAlaAlaGlyThrGlyGluAlaProValProGlyProG 67  
986 CFTGGAGCGGACGTCGAGGACAGAGGGAGGCCCGCTCCGCGCCCGGG 1035  
67 yAlaProLeuProAlaProLeuGlnValSerGluSerProTyrLeuH 84  
1036 AGCGCCGTGCCCCCGCCCGCGCTGCAGGTGTCGAAATCCTCGTGCTCC 1085  
84 isAlaProSerLeuLysThrSerCysGluTyrValSerLeuTyrHisGln 100  
1086 ATGCCCCATCTCTGAAGACGACCTGTGAATACGTAGCGCTCTACCCACG 1135  
101 ProAlaAlaMetMetGluAspSerAspSerAspPyrIleAsnValPr 117  
1136 CCTGCGCCCATGATGAGGAGCAGTGATTGAGAGCTACATCATGTGCC 1185  
117 oAla 118  
1186 TGCC 1189

seq\_name: qb\_htq11:AC023534

seq\_documentation\_block:  
LOCUS AC023534 211365 bp DNA HTG 02-MAR-2000  
DEFINITION Homo sapiens clone RP11-462N18, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 61  
unordered pieces.  
ACCESSION AC023534  
VERSION AC023534.2 GI:7143452  
KEYWORDS HTG: HTGS\_PHASE1.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 221365)  
Birken, B., Linton, L., Nusbaum, C. and Lander, E.  
Homo sapiens chromosome, clone RP11-462N18  
Unpublished  
2 (bases 1 to 221365)  
Birken, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
Anderson, S., Baldwin, J., Barna, N., Beda, F., Boguslavsky, L.,  
Boukanger, B., Brown, A., Burkett, G., Campopiano, A., Castle, A.,  
Choepel, J., Colangelo, M., Collins, S., Collumore, A., Cooke, P.,  
DeArellano, K., Dewar, K., Dodge, S., Domino, M., Doyle, M.,  
Fenster, J., Ferreira, P., FitzHugh, W., Forrest, C., Gage, D.,  
Galaun, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,  
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,  
Klein, J., Landers, T., Largoczek, J., Lehoczyk, J., Levine, R.,  
Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M.,  
McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrum, J.,  
Meneus, L., Minova, T., Miranda, C., Mienga, V., Morrow, J.,  
Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Olivart, T. M.,  
Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C.,  
Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S.,  
Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A.,

```
* 34913 36817: contig of 1905 bp in length
* 36818 36917: gap of 100 bp
* 36918 38575: contig of 1858 bp in length
* 38576 38675: gap of 100 bp
* 38676 41309: contig of 2634 bp in length
* 41310 41409: gap of 100 bp
* 41410 43344: contig of 1935 bp in length
* 43345 43444: gap of 100 bp
* 43445 45853: contig of 2409 bp in length
* 45854 45953: gap of 100 bp
* 45954 48434: contig of 2481 bp in length
* 48435 48534: gap of 100 bp
* 48535 51026: contig of 2492 bp in length
* 51027 51126: gap of 100 bp
* 51127 54369: contig of 3243 bp in length
* 54370 54469: gap of 100 bp
* 54470 56661: contig of 2192 bp in length
* 56662 56761: gap of 100 bp
* 56762 59097: contig of 2336 bp in length
* 59098 59197: gap of 100 bp
* 59198 61551: contig of 2354 bp in length
* 61552 61651: gap of 100 bp
* 61652 64257: contig of 2606 bp in length
* 64258 64357: gap of 100 bp
* 64358 67511: contig of 3154 bp in length
* 67512 67611: gap of 100 bp
* 67612 70329: contig of 2718 bp in length
* 70330 70429: gap of 100 bp
* 70430 73572: contig of 3143 bp in length
* 73573 73672: gap of 100 bp
* 73673 77203: contig of 3531 bp in length
* 77204 77303: gap of 100 bp
* 77304 80599: contig of 3296 bp in length
* 80600 80699: gap of 100 bp
* 80700 84046: contig of 3347 bp in length
* 84047 84146: gap of 100 bp
* 84147 88357: contig of 4211 bp in length
* 88358 88457: gap of 100 bp
* 88458 92454: contig of 3997 bp in length
* 92455 92554: gap of 100 bp
* 92555 96892: contig of 4338 bp in length
* 96893 96992: gap of 100 bp
* 96993 101334: contig of 4342 bp in length
* 101335 101434: gap of 100 bp
* 101435 105438: contig of 4004 bp in length
* 105439 105538: gap of 100 bp
* 105539 110413: contig of 4875 bp in length
* 110414 110513: gap of 100 bp
* 110514 115994: contig of 5481 bp in length
* 115995 116094: gap of 100 bp
* 116095 120447: contig of 4353 bp in length
* 120448 120547: gap of 100 bp
* 120548 124483: contig of 3936 bp in length
* 124484 124583: gap of 100 bp
* 124584 130249: contig of 5666 bp in length
* 130250 130349: gap of 100 bp
* 130350 135462: contig of 5113 bp in length
* 135463 135562: gap of 100 bp
* 135563 139847: contig of 4285 bp in length
* 139848 139947: gap of 100 bp
* 139948 146731: contig of 6784 bp in length
* 146732 146831: gap of 100 bp
* 146832 151469: contig of 4638 bp in length
* 151470 151569: gap of 100 bp
* 151570 159901: contig of 8332 bp in length
* 159902 160001: gap of 100 bp
* 160002 168308: contig of 8307 bp in length
* 168309 168408: gap of 100 bp
* 168409 176195: contig of 7787 bp in length
* 176196 176295: gap of 100 bp
* 176296 189303: contig of 13008 bp in length
* 189304 189403: gap of 100 bp
* 189404 202135: contig of 12732 bp in length
```

```
* 202136 202235: gap of 100 bp
* 202236 221365: contig of 19130 bp in length.
FEATURES
    source
        Location/Qualifiers
            1..221365
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="RP11-462N18"
                /clone_lib="RPC1-11 Human Male BAC"
            1..1108
                /note="assembly_fragment"
            1209..2292
                /note="assembly_fragment"
            2393..3673
                /note="assembly_fragment"
            3774..4947
                /note="assembly_fragment"
            5048..6104
                /note="assembly_fragment"
            6205..7259
                /note="assembly_fragment"
            7360..8667
                /note="assembly_fragment"

alignment_scores:
    Quality: 325.00      Length: 144
    Ratio: 4.333        Gaps: 1
    Percent Similarity: 52.083      Percent Identity: 50.000

alignment_block:
    US-09-135-238B-2_COPY_273_390 x AC023534 ..
    Align seg 1/1 to: AC023534 from: 1 to: 221365

      8 LysAlaLeuSerArgAlaArgAlaValArgMetArgAlaLeu 24
      :::::::::::::::::::::::::::::::::::::::::::
32391 GAAGCCCTCTCCAGGCGGGCGCGGCGGCTGGCGGTGAGGATCGCGCCCT 32440
      |||
24 uGluSerSerGlnArgProArgGlySerProArgProArgSerGlnAsnA 41
      |||
32441 GGAGAGCTCCAGAGGCGCGCGGGGTGCGCGGCGGCGGCTCCCAAAACA 32490
      |||
41 snllyrSerAlaCysProArgArgAlaArgGlyAlaAspAlaAla... 56
      |||
32491 ACATCTACAGCGCTCCCGCGGCGCGCTCGTGTGAGCGGACGCTCGAGGT 32540
      |||
56 ..... 56

32541 GAGCCCGCGGGCGGCGGAGTAGCGGAAAGAGTGCGCTGGCTCGG 32590
      |||
56 ..... 56

32591 GGAAGCGCGGGCGGCGGCGGCGGCTCCAGGTGCCAGGGTCCACACA 32640
      |||
56 ..... 56

32641 GGTGTGAGCTGGAGTGGGGGGGGGTGGAGCTGGAGTAGTAGCATGAC 32690
      |||
56 ..... 56

32691 GGTGGGGCGGAGCGCTCGGCGCGCGGCTCACCTCGCGCTGTGCGCCT 32740
      |||
57 .GlyThrGlyGluAlaProValProGlyProGlyAlaProLeuProProA 73
      |||
32741 AGGCACAGGGAGGGCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 32790
      |||
73 laProLeuGlnValSerGluSerProTrpLeu 83
      |||
32791 CCCCGGTGCAGGTAAAGCTCCGCTCTCGCCCTG 32822

seq_name: gb.htg20:AC010538

seq_documentation_block:
    LOCUS AC010538 249227 bp DNA HTG
    DEFINITION Homo sapiens chromosome 16 clone RP11-368I7, WORKING DRAFT
```

28-MAR-2000

SEQUENCE, 90 unordered pieces.  
AC010538  
VERSION AC010538.3 GI:7331338  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 249227)  
DOE Joint Genome Institute.  
TITLE Sequencing of Human Chromosome 16  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 249227)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
On Mar 28, 2000 this sequence version replaced gi:6758794.  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: <http://www.jgi.doe.gov>  
-----Summary Statistics  
Consensus quality: 189807 bases at least Q40  
Consensus quality: 22221 bases at least Q30  
Consensus quality: 230312 bases at least Q20  
Estimated insert size: 249227; sum-of-contigs estimation  
Estimated insert size: 245950; agarose-fp estimation  
Quality coverage: 3.16x in Q20 bases; agarose-fp estimation  
Quality coverage: 3.12x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 90 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 1060: contig of 1060 bp in length  
\* gap of unknown length  
\* 1061 2237: contig of 1177 bp in length  
\* gap of unknown length  
\* 2238 3270: contig of 1033 bp in length  
\* gap of unknown length  
\* 3271 4550: contig of 1280 bp in length  
\* gap of unknown length  
\* 4551 5647: contig of 1097 bp in length  
\* gap of unknown length  
\* 5648 6804: contig of 1157 bp in length  
\* gap of unknown length  
\* 6805 7807: contig of 1003 bp in length  
\* gap of unknown length  
\* 7808 9086: contig of 1279 bp in length  
\* gap of unknown length  
\* 9087 10101: contig of 1015 bp in length  
\* gap of unknown length  
\* 10102 11251: contig of 1150 bp in length  
\* gap of unknown length  
\* 11252 12446: contig of 1195 bp in length  
\* gap of unknown length  
\* 12447 13557: contig of 1111 bp in length  
\* gap of unknown length  
\* 13558 14576: contig of 1019 bp in length  
\* gap of unknown length  
\* 14577 15705: contig of 1129 bp in length  
\* gap of unknown length  
\* 15706 17020: contig of 1315 bp in length  
\* gap of unknown length  
\* 17021 18385: contig of 1365 bp in length  
\* gap of unknown length  
\* 18386 19680: contig of 1295 bp in length  
\* gap of unknown length  
\* 19681 20882: contig of 1202 bp in length  
\* gap of unknown length  
\* 20883 21954: contig of 1072 bp in length  
\* gap of unknown length  
\* 21955 23689: contig of 1735 bp in length  
\* gap of unknown length  
\* 23690 24962: contig of 1273 bp in length  
\* gap of unknown length  
\* 24963 26319: contig of 1357 bp in length  
\* gap of unknown length  
\* 26320 27582: contig of 1263 bp in length  
\* gap of unknown length  
\* 27583 28849: contig of 1267 bp in length  
\* gap of unknown length  
\* 28850 29861: contig of 1012 bp in length  
\* gap of unknown length  
\* 29862 31003: contig of 1142 bp in length  
\* gap of unknown length  
\* 31004 32367: contig of 1364 bp in length  
\* gap of unknown length  
\* 32368 33785: contig of 1418 bp in length  
\* gap of unknown length  
\* 33786 34834: contig of 1049 bp in length  
\* gap of unknown length  
\* 34835 35941: contig of 1107 bp in length  
\* gap of unknown length  
\* 35942 37397: contig of 1456 bp in length  
\* gap of unknown length  
\* 37398 38403: contig of 1006 bp in length  
\* gap of unknown length  
\* 38404 39501: contig of 1098 bp in length  
\* gap of unknown length  
\* 39502 40724: contig of 1223 bp in length  
\* gap of unknown length  
\* 40725 41991: contig of 1267 bp in length  
\* gap of unknown length  
\* 41992 43426: contig of 1435 bp in length  
\* gap of unknown length  
\* 43427 45236: contig of 1810 bp in length  
\* gap of unknown length  
\* 45237 46594: contig of 1458 bp in length  
\* gap of unknown length  
\* 46595 47746: contig of 1052 bp in length  
\* gap of unknown length  
\* 47747 49191: contig of 1445 bp in length  
\* gap of unknown length  
\* 49192 51044: contig of 1853 bp in length  
\* gap of unknown length  
\* 51045 52378: contig of 1334 bp in length  
\* gap of unknown length  
\* 52379 53505: contig of 1127 bp in length  
\* gap of unknown length  
\* 53506 55286: contig of 1781 bp in length  
\* gap of unknown length  
\* 55287 56793: contig of 1507 bp in length  
\* gap of unknown length  
\* 56794 58628: contig of 1855 bp in length  
\* gap of unknown length  
\* 58629 59765: contig of 1137 bp in length  
\* gap of unknown length  
\* 59766 61158: contig of 1393 bp in length  
\* gap of unknown length  
\* 61159 62688: contig of 1530 bp in length  
\* gap of unknown length  
\* 62689 64788: contig of 2100 bp in length  
\* gap of unknown length  
\* 64789 66536: contig of 1748 bp in length  
\* gap of unknown length  
\* 66537 70463: contig of 3927 bp in length  
\* gap of unknown length  
\* 70464 72740: contig of 2277 bp in length  
\* gap of unknown length  
\* 72741 75083: contig of 2343 bp in length  
\* gap of unknown length

```

* 75084 gap of unknown length
* 77658: contig of 2575 bp in length
* 77659 gap of unknown length
* 79607: contig of 1949 bp in length
* 79608 gap of unknown length
* 81867: contig of 2260 bp in length
* 81868 gap of unknown length
* 84131: contig of 2264 bp in length
* 84132 gap of unknown length
* 86256: contig of 2125 bp in length
* 86257 gap of unknown length
* 88488: contig of 2232 bp in length
* 88489 gap of unknown length
* 90327: contig of 1839 bp in length
* 90328 gap of unknown length
* 92118: contig of 1791 bp in length
* 92119 gap of unknown length
* 94665: contig of 2547 bp in length
* 94666 gap of unknown length
* 97095: contig of 2430 bp in length
* 97096 gap of unknown length
* 99715: contig of 2620 bp in length
* 99716 gap of unknown length
* 102136: contig of 2421 bp in length
* 102137 gap of unknown length
* 104714: contig of 2578 bp in length
* 104715 gap of unknown length
* 107130: contig of 2416 bp in length
* 107131 gap of unknown length
* 109573: contig of 2443 bp in length
* 109574 gap of unknown length
* 112859: contig of 3286 bp in length
* 112860 gap of unknown length
* 115291: contig of 2432 bp in length
* 115292 gap of unknown length
* 117865: contig of 2574 bp in length
* 117866 gap of unknown length
* 121526: contig of 3661 bp in length
* 121527 gap of unknown length
* 125987: contig of 4461 bp in length
* 125988 gap of unknown length
* 132407: contig of 8420 bp in length
* 132408 gap of unknown length
* 137033: contig of 4626 bp in length
* 137034 gap of unknown length
* 141135: contig of 4102 bp in length
* 141136 gap of unknown length
* 146850: contig of 5715 bp in length
* 146851 gap of unknown length
* 152336: contig of 5486 bp in length
* 152337 gap of unknown length
* 159908: contig of 7572 bp in length

alignment_scores:
  Quality: 107.00      Length: 84
  Ratio: 2.378        Gaps: 4
  Percent similarity: 53.571 Percent Identity: 38.095

alignment_block:
US-09-135-238b-2_COPY_273_390 x AC010538/rev ..

Align seg 1/1 to reverse of: AC010538 from: 1 to: 249227

32 GlySerProArgProArgSerGlnAsnIleTyrSerAlaCysProAr 48
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
112713 GGAACCCCGCGCCGCGACACCCAGACCCAGCGCTACCCAGCCCG 112664
48 q.....:.....:.....:.....:.....:.....:.....
      |.....:.....:.....:.....:.....:.....:.....
112663 GCACCCCGCGCGCCGACCCCGACAGGGCTCGACCGCCGCGAC 112614
55 .....AlaAlaGlyThrGlyGlu.....:.....:.....:.....
      ||| ||||| ||| :|||:|||||:|||||:|||||:|||||

```

```

112613 CGCCAGCCCGCGAGAACCCCGAGACCGTGGCTCAGCCCGCCAGCTCCC 112564
67 GlyAlaProLeuProAlaProLeuGlnValSerGluSerProTrrLe 83
|||||:|||||:|||||:|||||:|||||:|||||:|||||
112563 CCAGCCCGCGCTCCCGCCAGCCCGCTCTCCAGCCCGCCAGTGGCT 112514
83 uHisAlaProSerLeuLysThrSerCysGluTyrValSerLeuTyrHisG 100
|||||:|||||:|||||:|||||:|||||:|||||
112513 AGGTATCCCTAAG.....CACCTGGCTCAGGTGAGCACTCACCATC 112473
100 ln 100
||
112472 AA 112471

seq_name: gb_bal:SCI35

seq_documentation_block:
LOCUS SCI35 40909 bp DNA BCT 09-SEP-1998
DEFINITION Streptomyces coelicolor cosmid I35.
ACCESSION AL031541
VERSION AL031541.1 GI:3581833
KEYWORDS 3-hydroxyacyl-coa dehydrogenase; 50S ribosomal protein L20; 50S
ribosomal protein L35; arabinosidase; argC; argJ; ATC initiation
codon; ATP/GTP binding protein; dehydrogenase; glutamate
N-acetyltransferase; glutamine synthetase; infC; IS117; mutT
domain; N-acetyl-gamma-glutamyl-phosphate reductase;
phenylalanyl-tRNA synthetase; pheS; pheT; rplT; rpmI; rRNA
methylase; secreted serine protease; transcriptional regulator;
translation initiation factor IF-3; transposase; TTA leu codon;
two-component sensor.
SOURCE Streptomyces coelicolor A3(2).
ORGANISM Streptomyces coelicolor A3(2).
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 40909)
AUTHORS Oliver, K. and Harris, D.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 40909)
AUTHORS Parkhill, J., Barrell, B. G. and Rajandream, M. A.
TITLE Direct Submission
JOURNAL Submitted (08-SEP-1998) Streptomyces coelicolor sequencing project,
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
Colney, Norwich, Norfolk NR4 7UH, UK
REFERENCE 3 (bases 1 to 40909)
AUTHORS Redenbach, M., Kieser, H. M., Denapaite, D., Eichner, A., Cullum, J.,
Kinashi, H. and Hopwood, D. A.
TITLE A set of ordered cosmids and a detailed genetic and physical map
for the 8 Mb Streptomyces coelicolor A3(2) chromosome
JOURNAL Mol. Microbiol. 21 (1), 77-96 (1996)
MEDLINE 97000351
COMMENT
Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded
by the BBSRC.
Details of S. coelicolor sequencing at the Sanger Centre are
available on the World Wide Web.
(URL: http://www.sanger.ac.uk/Projects/S\_coelicolor/) CDS are
numbered using the following system eg SC787.01c. SC (S.
coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
strand).
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous. The length
in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for
CDS which show significant similarity to other CDS in the database.
The position of possible ribosome binding site sequences are given
where these have been used to deduce the initiation codon. Gene
prediction is based on positional base preference in codons using a
specially developed Hidden Markov Model (Krogn et al., Nucleic
Acids Research, 22(22):4768-4778(1994)) and the FramePlot program
of Bibb et al., Gene 30:157-66(1984) as implemented at
http://www.nih.go.jp/

```

**CAUTION:** We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtc or att) which is preceded by an upstream ribosome binding site sequence (optimally 5-10bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

**IMPORTANT:** This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid I35 lies between t10 and t41 on the AseI-I genomic restriction fragment.

FEATURES  
SOU

source

```

i. 4.03039
  /organism="Streptomyces coelicolor A3(2)"
  /strain="A3(2)"
  /db_xref="taxon:100226"
  /clone="cosmid I35"
  complement(1..552)
  /gene="argJ"
  complement(<1..552)

```

/note="SC135\_01c, argJ, probable glutamate N-acetyltransferase, partial CDS, len: >184 aa; similar to many e.g. ARGJ\_CORGL glutamate N-acetyltransferase (ornithine acetyltransferase) (EC 2.1.1.35) (388 aa), fasta scores; opt: 614 z-score: 909.0 (E): 0, 55.4% identity in 184 aa overlap. Referred to as 'argB' on genetic map"

```

/codon_table=1
/transl_start=1
/product="putative glutamate N-acetyltransferase"
/protein_id="CAA20790.1"
/db_xref="GI:3581834"
/db_xref="SPRMBL:O88040"
/translation="MSVTRAGFTFRAGITAGIKESQSPDLAVNTGPPESRAGVFTIS
NRVKAAPVIMSEVQSKSEVAVLVNMSGANACTGQPDTHATPKADVLCGAG
EVKAVSTGLGILPMDKLPLPEVNAAGLSEHGEGAKAALIKTTDTHVTKISVTRDGG
WTGGMARGAGMLAPGLATIMLVIT
complement(549..1577)

```

/gene="argC"  
 complement(549. .1577)  
 /gene="argC"  
 /note="SC135.02c, argC, N-acetyl-gamma-glutamyl-phosphate  
 reductase, len: 342 aa; identical to ARGCS49CO  
 N-acetyl-gamma-glutamyl-phosphate reductase (EC 1.2.1.38)  
 (249 aa). Contains PS01224  
 N-acetyl-gamma-glutamyl-phosphate reductase active site.  
 Referred to as 'argAV' on genetic map."

```

/codon_start=1
/trans_table=1
/product="N-acetyl-gamma-glutamyl-phosphate reductase"
/protein_id="CAA30791.1"
/db_xref="GI:3581835"
/translation="MAVRAAGASGAYAGGELLRLTLTPEVIGALGTGNSAQRLG
ALOPFLIADLRVLEATPEVLGGHDFVFLALPHQSAVAEQDVLVIVMDGDFR
LKDAGWEFYFGSHATGYKPLPGLGARAALGSKRIAPCGCTYATVSLAFPAYA
ASLAEPEVIAASTGSGACKAPLPHSGSVMGSMGSPGCGGHHRTPEMIGNLGA
AGEVPTVSTPTPLAPWRPGIATKATKALKEVTAESYARA EKALADSPVFVLLPQCG
AGVATL SVKSGSVKGVHVAHDAAGRTALATSTGNTACTGASAVOSMVAALGIDETTG

```

```

LTTIGVAP"
complement(560..563)
/gene="argC"
/note="possible RBS upstream of argJ"
complement(831..1877)
/note="99.8% match to SCARGC X66783 S.coelicolor argC gene"
/real="1"

```

complement(1107..1157)  
/gene="argC"  
/note="PS01224 N-acetyl-gamma-glutamyl-phosphate reductase  
active site"  
complement(1585..1590)  
/note="possible RBS upstream of argC"  
1611..2174

```

/genes="SCI35.03"
1611. .2174
/genes="SCI35.03"
/notes="SCI35.03, unknown, len: 187 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein SCI35.03"
/protein_id="CAA20792.1"
/db_xref="GI:3581836"
/db_xref="SPTREMBL:O88041"
/translation="MPHEVYAMCVGTSSRRRRRLACRFPTPTGVGVAVYAMIVELV
GLRPQRAALVSLCALALPLHFAVGNDFVROGOTAVTEGFVCEWDGDCPGAGWRL
ADSLGRGIDGLDGLTEELVEDIPLFGGRWADVADRGTLRHAVAQYTVAPVGAAL
VVIARRRAYNQETALMKRYRTTP"
complement(2164. .2688)
/genes="SCI35.04c"
complement(2164. .2688)
/genes="SCI35.04c"
/notes="SCI35.04c, ATP/GTP binding protein, len: 174 aa;
similar to TMREB_BACSU tunicamycin resistance protein (197
aa), fasta scores, opt: 282 z-score: 488.2 E(): 6.1e-20,
39.5% identity in 119 aa overlap. Contains PS00017
ATP/GTP-binding site motif A (P-loop)"
/codon_start=1
/transl_table=11
/product="ATP/GTP binding protein"
/protein_id="CAA20793.1"
/db_xref="GI:3581837"
/db_xref="SPTREMBL:O88042"
/translation="MIWVLNTHGAGKTTTGALYQRLLPDSRVFDAEKVGFELMDLITP
GLPTDFNQHWAPPLVFAVTVLDITGGTVLPMPTVLVLEIWRRETAGSLAAGHP
VRIHFLVHAQDETILRGIRAGTVTLGPDSPERLYLEPYAAEAARTLWHAEEVVDTHLT
PAEAALRIAPAVRA"
complement(2647. .2670)
misc feature

```

```

gene
complement(2751..5225)
/gene="SCI35_05c"
CDS
complement(2751..5225)
/gene="SCI35_05c"
/note="SCI35_05c, probable secreted arabinosidase, len: 824 aa; similar to e.g. TR:Q59218 (EMBL:U015178) arabinosidase [EC 3.2.1.55] (Bacteroides ovatus) (560 aa), fasta scores: opt: 1204 z-score: 1030.7 E(): 0, 41.4% identity in 601 aa overlap"
/codon_start=1
/transl_table=11
/product="putative secreted arabinosidase"
/protein_id="CAA20794.1"
/db_xref="GI:3581838"
/db_xref="SPTREMBL:O88043"
/translations="MSRLRWLGUSVITALLTAAALLSPAAEAADVITIVDPDSRGPAIDTWHGVYFEDINRAADGLYAEIVONRSEFYEADNASYPTLTAWADGTAEVVNDGGLERNRNTLSLAKASGVTVINRGVIRGEKGYDFSVMSAGHTLVLVITDAAGTGLKDLAEIKIALGPGYGFYFRFADIGAMPVPVATYGGQNKAVDDALLHOPEGTANANQVYFIEYSGLYEYFRFADIGAMPVPVATYGGQNKAVDDALLKRIHQDTLDITIEFANGPTSGKQKRAEKGKPFHLTHLEVENEENLPKEFFARFQDFTFLAELKYVTVVNSGSDPGATFTAWOLNADGVMDVEHYNSPQWFLQNDYRTSDNRNGPKVLYGASQGNANKNLALEAFMTGLERNADVVKLASVPIGLSNEYVQWSPDLLIWFNNHWSGANSYEVOKLFPMNNTGDRVVPSTAGTGPSGIPVGGVGLSTWATSAAYDDVKVTSDFGETLLSDFGDASGRWAHSGAGSVTVDQGTVGTDAEAENTHVTAGDPAWHIDYDHLVARTKSGEGLTVAFGVKDTGNTYWNNGWNNTOSASQASDDGKSTLTLSKAGSTIEGTRAVYDVKVGRQVTLYLDEGEWEGFTDDKPAEPRTQVITVSDRTGELIVKVVNAQDTPAARTAVDTLGGAKVASRAAVTTLAADDQAVNTGTDPVTPVTSFTSGASEFTYTPPANSVTEFLRIQR"
/note="possible RBS upstream of SCI35_05c"
5337..5340
/note="possible RBS upstream of SCI35_06"
5362..5820
gene
RBS
RBS

```

alignment scores:







AUTHORS Ricke,D.O.  
 TITLE Large Scale Sequence Analysis and Annotation with the Sequence  
 JOURNAL Comparison Analysis (SCAN) System  
 REFERENCE Unpublished  
 AUTHORS 3 (bases 1 to 154218)  
 Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D.,  
 Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L.,  
 Bryant,J., Tesner,J., Melnick,L., Longmire,J., White,S., Tatum,O.,  
 Campbell,C., Fawcett,J., Maltbie,M., Bussod,M., Sutherland,R.,  
 McMurry,K., Han,C. and Deaven,L.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-JAN-2000) Center for Human Genome Studies, DOE Joint  
 Genome Institute, Los Alamos National Laboratory, MS M888, Los  
 Alamos, NM 87545, USA  
 FEATURES  
 source  
 1. .154218  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="16"  
 /clone="RPCI-11.127120"  
 complement(41. .330)  
 rpt\_family="Alu"  
 575. .882  
 /rpt\_family="Alu"  
 1414. .1712  
 /rpt\_family="Alu"  
 1595. .1862  
 /rpt\_family="Alu"  
 1871. .2190  
 /rpt\_family="Alu"  
 2176. .2372  
 /rpt\_family="L1"  
 2498. .2795  
 /rpt\_family="Alu"  
 3187. .3504  
 /rpt\_family="Alu"  
 complement(3571. .3890)  
 /rpt\_family="Alu"  
 complement(3859. .4163)  
 4198. .4503  
 /rpt\_family="Alu"  
 4826. .5155  
 /rpt\_family="Alu"  
 5749. .6045  
 /rpt\_family="Alu"  
 complement(6848. .6983)  
 7473. .7776  
 /rpt\_family="Alu"  
 7931. .8385  
 /rpt\_family="Alu"  
 8808. .9139  
 /rpt\_family="Alu"  
 9394. .9682  
 /rpt\_family="Alu"  
 9515. .9812  
 /rpt\_family="Alu"  
 9702. .9987  
 /rpt\_family="Alu"  
 9965. .10332  
 /rpt\_family="Alu"  
 complement(10403. .10835)  
 /rpt\_family="Alu"  
 complement(10801. .10977)  
 11580. .11780  
 /rpt\_family="Alu"  
 /rpt\_family="MER42"  
 11803. .12101  
 /rpt\_family="Alu"  
 12282. .12576  
 /rpt\_family="Alu"  
 12643. .13073

---

repeat\_region /rpt\_family="Alu"  
 12923. .13237  
 /rpt\_family="Alu"  
 13420. .13725  
 /rpt\_family="Alu"  
 13761. .14148  
 /rpt\_family="Alu"  
 14072. .14371  
 /rpt\_family="Alu"  
 complement(14401. .14696)  
 /rpt\_family="Alu"  
 15061. .15366  
 /rpt\_family="Alu"  
 15383. .15706  
 /rpt\_family="Alu"  
 complement(16083. .16185)  
 /rpt\_family="L1"  
 complement(16192. .16347)  
 /rpt\_family="Alu"  
 16359. .16831  
 /rpt\_family="Alu"  
 complement(18420. .18718)  
 /rpt\_family="Alu"  
 complement(18518. .18917)  
 /rpt\_family="Alu"  
 complement(18754. .19036)  
 /rpt\_family="Alu"  
 complement(19456. .19559)  
 /note="GRAIL 2 excellent exon, frame 1"  
 20239. .20397  
 /rpt\_family="MER20"  
 complement(20808. .21104)  
 /rpt\_family="Alu"  
 complement(22116. .22443)  
 /rpt\_family="Alu"  
 complement(22789. .22948)  
 /note="GRAIL 2 excellent exon, frame 1"  
 23346. .23646  
 /rpt\_family="Alu"  
 complement(24046. .24341)  
 /rpt\_family="Alu"  
 25205. .25461  
 /rpt\_family="Alu"  
 25605. .25862  
 /rpt\_family="L1"  
 25914. .25981  
 /rpt\_family="Alu"  
 complement(26044. .26355)  
 /rpt\_family="Alu"  
 28437. .28732  
 /rpt\_family="Alu"  
 30597. .30772  
 /note="GRAIL 2 excellent exon, frame 2"  
 31300. .31655  
 /rpt\_family="Alu"  
 32059. .32124  
 /rpt\_family="Alu"  
 complement(32302. .32492)  
 /rpt\_family="Alu"  
 32507. .32839  
 /rpt\_family="Alu"  
 33649. .34040  
 /rpt\_family="Alu"  
 complement(34202. .34278)  
 /rpt\_family="L1"  
 34279. .34556  
 /rpt\_family="Alu"  
 35017. .35495  
 /rpt\_family="Alu"  
 complement(36639. .36991)  
 /rpt\_family="Alu"  
 37667. .37808  
 /note="GRAIL 2 excellent exon, frame 0"

```
repeat_region complement(37920..37961)
repeat_region /rpt_family="MIR"
repeat_region 38430..38745
repeat_region /rpt_family="Alu"
repeat_region 38789..38899
repeat_region /rpt_family="MER3"
repeat_region complement(40939..41089)
repeat_region /rpt_family="Alu"
repeat_region 42294..42589
repeat_region /rpt_family="Alu"
repeat_region complement(43093..43396)
repeat_region /rpt_family="Alu"
repeat_region complement(43274..43695)
repeat_region /rpt_family="Alu"
repeat_region 44405..44482
repeat_region /rpt_family="MLTIC"
repeat_region 44515..44693
repeat_region /rpt_family="Alu"
repeat_region 45816..46254
repeat_region /rpt_family="Alu"
misc_feature complement(46300..46347)
repeat_region /note="GRAU 2 excellent exon, frame 1"
repeat_region 46426..46596
repeat_region /rpt_family="MER20"
repeat_region 47139..47448
repeat_region /rpt_family="Alu"
repeat_region 47535..47943
repeat_region /rpt_family="THE1"
repeat_region complement(47939..48377)
repeat_region /rpt_family="Alu"
repeat_region 48431..48778
repeat_region /rpt_family="L1"
repeat_region 48877..49260
repeat_region /rpt_family="Alu"
repeat_region complement(49310..49588)
repeat_region /rpt_family="Alu"

alignment_scores:
  Quality: 104.50      Length: 96
  Ratio: 2.133        Gaps: 4
  Percent Similarity: 51.042  Percent Identity: 36.458

alignment_block:
US-09-135-238B-2_COPY_273_390 x AC020663 ..
Align seg 1/1 to: AC020663 from: 1 to: 154218

      2 ArgAlaValGluArgArg..... 7
140217 CGCGCGCGGAGCGGAGACTCCCGCTCTCCCTCCCTTCGGCTC 140266
      8 .....LysAlaLeuSerArgArgAlaArgAlaVal 19
140267 GTGACACGAGCGCGCGGCTGAGCGCGCGCGCGCGCGGTGC 140316
      19 aAlaArgMetArgAlaLeuGluSerSerGlnArgProArgGlySerProArg 35
140317 ACCGGCTGAGCGGAGAGGAGC.....CGGCTCGCGGCTCGCCCGC 140360
      36 ProArgSerGlnAsnAsnIleTyrSerAlaCysPro.....ArgAr 49
140361 CCCCCG.....GTCCTGGACTCCGCGCCCTCCCTCTCGGCG 140398
      49 gAlaArgGlyAlaAspAlaGlyThrGlyGluAlaProValProGlyP 66
140399 CTTCCGTTACCGCGTGTGTCGCGCGCGCGCGCGCGGCGCCATCCGAGC 140448
      66 roGlyAlaProLeuProAlaProLeuInValSer 78
140449 CCGAGCGCTGTCGCGCGCGTTCGCGAGCGCGCAGTGAGC 140486
seq_name: gb_hg22:AC027687

seq_documentation_block:
LOCUS AC027687 245472 bp DNA HTG 01-APR-2000
DEFINITION Homo sapiens chromosome 16 clone RP11-6114, WORKING DRAFT SEQUENCE,
50 unordered pieces.
ACCESSION AC027687
VERSION AC027687.1 GI:7381751
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 245472)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 245472)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (01-APR-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----Summary Statistics
Consensus quality: 177301 bases at least Q40
Consensus quality: 204545 bases at least Q30
Consensus quality: 220813 bases at least Q20
Estimated insert size: 245472; sum-of-contigs estimation
Estimated insert size: 169000; agarose-fp estimation
Quality coverage: 7.37x in Q20 bases; agarose-fp estimation
Quality coverage: 5.08x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 50 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1298: contig of 1298 bp in length
* gap of unknown length
* 1299 2303: contig of 1005 bp in length
* gap of unknown length
* 2304 3390: contig of 1087 bp in length
* gap of unknown length
* 3391 4442: contig of 1052 bp in length
* gap of unknown length
* 4443 5615: contig of 1173 bp in length
* gap of unknown length
* 5616 6822: contig of 1207 bp in length
* gap of unknown length
* 6823 8057: contig of 1235 bp in length
* gap of unknown length
* 8058 9328: contig of 1271 bp in length
* gap of unknown length
* 9329 11175: contig of 1847 bp in length
* gap of unknown length
* 11176 12414: contig of 1239 bp in length
* gap of unknown length
* 12415 13648: contig of 1234 bp in length
* gap of unknown length
* 13649 14938: contig of 1290 bp in length
* gap of unknown length
* 14939 16002: contig of 1064 bp in length
* gap of unknown length
* 16003 17497: contig of 1495 bp in length
* gap of unknown length
* 17498 19150: contig of 1653 bp in length
* gap of unknown length
* 19151 20166: contig of 1016 bp in length
* gap of unknown length
```

\* 20167 21607: contig of 1441 bp in length  
\* gap of unknown length  
\* 21608 22911: contig of 1304 bp in length  
\* gap of unknown length  
\* 22912 24331: contig of 1420 bp in length  
\* gap of unknown length  
\* 24332 26162: contig of 1831 bp in length  
\* gap of unknown length  
\* 26163 28064: contig of 1902 bp in length  
\* gap of unknown length  
\* 28065 29929: contig of 1865 bp in length  
\* gap of unknown length  
\* 29930 31543: contig of 1614 bp in length  
\* gap of unknown length  
\* 31544 33044: contig of 1501 bp in length  
\* gap of unknown length  
\* 33045 35128: contig of 2084 bp in length  
\* gap of unknown length  
\* 35129 37580: contig of 2452 bp in length  
\* gap of unknown length  
\* 37581 38859: contig of 1279 bp in length  
\* gap of unknown length  
\* 38860 40904: contig of 2045 bp in length  
\* gap of unknown length  
\* 40905 43040: contig of 2136 bp in length  
\* gap of unknown length  
\* 43041 44561: contig of 1521 bp in length  
\* gap of unknown length  
\* 44562 46554: contig of 1993 bp in length  
\* gap of unknown length  
\* 46555 48508: contig of 1954 bp in length  
\* gap of unknown length  
\* 48509 49737: contig of 1229 bp in length  
\* gap of unknown length  
\* 49738 51391: contig of 1654 bp in length  
\* gap of unknown length  
\* 51392 52987: contig of 1596 bp in length  
\* gap of unknown length  
\* 52988 55258: contig of 2271 bp in length  
\* gap of unknown length  
\* 55259 58043: contig of 2785 bp in length  
\* gap of unknown length  
\* 58044 59391: contig of 1348 bp in length  
\* gap of unknown length  
\* 59392 62496: contig of 3105 bp in length  
\* gap of unknown length  
\* 62497 65618: contig of 3122 bp in length  
\* gap of unknown length  
\* 65619 67648: contig of 2030 bp in length  
\* gap of unknown length  
\* 67649 70075: contig of 2427 bp in length  
\* gap of unknown length  
\* 70076 73261: contig of 3186 bp in length  
\* gap of unknown length  
\* 73262 76990: contig of 3629 bp in length  
\* gap of unknown length  
\* 76991 80022: contig of 3132 bp in length  
\* gap of unknown length  
\* 80023 89215: contig of 9193 bp in length  
\* gap of unknown length  
\* 89216 98779: contig of 9564 bp in length  
\* gap of unknown length  
\* 98780 132048: contig of 33269 bp in length  
\* gap of unknown length  
\* 132049 190353: contig of 58305 bp in length  
\* gap of unknown length  
\* 190354 245472: contig of 55119 bp in length.  
Location/Qualifiers  
1. .245472  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="RP11-61L4"

FEATURES  
source

BASE COUNT 64057 a 60702 c 58925 g 61440 t 348 others  
ORIGIN  
  
alignment\_scores:  
Quality: 104.50 Length: 96  
Ratio: 2.133 Gaps: 4  
Percent Similarity: 51.042 Percent Identity: 36.458  
  
alignment\_block:  
US-09-135-238B-2\_COPY\_273\_390 x AC027687 ..  
Align seg 1/1 to: AC027687 from: 1 to: 245472  
  
2 ArgAlaValGluArgArg..... 7  
||||| |||||||  
172086 CGCGCGCGGAGCGAGACTCCCGGCTCTCCCTCCCTCGGTC 172135  
  
8 .....LysAlaLeuSerArgAlaArgAlaArgLeuAlav 19  
172136 GTGACAACGACGCCCGGCTCTGAGCGCGCGCGGCGACGTGCG 172185  
  
19 alArgMetArgAlaLeuGluSerSerGlnArgProArgGlySerProArg 35  
||||| ||| ||||||| ||||||| |||||||  
172186 ACCGGCTGAGCGGAGAGGAGC.....CGGCTCGGGGCTCGCCCGC 172229  
  
36 ProArgSerGlnAsnAsnIleTyrSerAlaCysPro.....ArgAr 49  
||||| : : : : : |||||  
172230 CCCCCG.....GTCTTGACTCCGCGCCCTCTCTCGCG 172267  
  
49 gAlaArgGlyAlaAspAlaAlaGlyThrGlyGluAlaProValProGlyP 66  
||||| ||| : : : : : |||||  
172268 CTTCCGTACGCCCTTGTCGCGCGCGCGCGGCCATCCCGAGC 172317  
  
66 roGlyAlaProLeuProAlaProLeuGlnValSer 78  
|| ||| : : : |||||  
172318 CGAGGCGCTGTGTCGCCCTCGCAGCGCCAGTGAGC 172355  
  
seq\_name: gb\_bal:SCF43A

seq\_documentation\_block:

LOCUS SCF43A 35437 bp DNA BCT 13-JUL-1999  
DEFINITION Streptomyces coelicolor cosmid F43A.  
ACCESSION AL096837  
VERSION AL096837.1 GI:5459214  
KEYWORDS

ABC transporter; alkaline phosphatase; ATP-binding protein;  
deoxyribodipyrimidine photolyase; ferredoxin; gntR-family;  
iron-sulfur protein; isomerase; lyase; oxidoreductase;  
penicillin-binding protein; phr; rhamnose kinase; serine protease;  
threonine dehydratase; transcriptional regulator; transmembrane  
transport protein; transport protein.

SOURCE  
ORGANISM

Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

Actinomycetales; Streptomycineae; Streptomyces; Streptomyces.

REFERENCE

AUTHORS

Redenbach, M., Kieser, H.M., Denapaite, D., Eichner, A., Cullum, J.,

Kinashi, H. and Hopwood, D.A.

A set of ordered cosmids and a detailed genetic and physical map

for the 8 Mb Streptomyces coelicolor A3(2) chromosome

Mol. Microbiol. 21 (1), 77-96 (1996)

97000351

2 (bases 1 to 35437)

Seeger, K. and Harris, D.

Unpublished

JOURNAL

REFERENCE

3 (bases 1 to 35437)

James, K.D., Parkhill, J., Barrell, B.G. and Rajandream, M.A.

Direct Submission

JOURNAL

Submitted (09-JUL-1999)

Streptomyces coelicolor sequencing project,  
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge  
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.  
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,  
Colney, Norwich, Norfolk NR4 7UH, UK

## COMMENT

Notes:  
 Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics  
 Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.  
 (URL: <http://www.sanger.ac.uk/Projects/S.coelicolor/>)  
 CDS are numbered using the following system eg SC787.01c, SC (S. coelicolor), 787 (cosmid name), .01 (first CDS), c (complementary strand).  
 The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.  
 The length in codons is given for each CDS.  
 Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database.  
 The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon.  
 Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.gov/jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.  
 IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.  
 Cosmid F43A lies between cosmids F43 and M2 on the AseI-F region of the chromosome.  
 Location/Qualifiers  
 1..35437  
 /organism="Streptomyces coelicolor A3(2)"  
 /strain="A3(2)"  
 /db\_xref="taxon:100226"  
 /clone="cosmid F43A"  
 complement(1..312)  
 /gene="SCF43A.01c"  
 complement(31..312)  
 /gene="SCF43A.01c"  
 /note="SCF43A.01c, probable ABC transporter ATP-binding protein, partial CDS, len: >104 aa; similar to many e.g. SW:ARAG\_ECOLI (EMBL:X06091), araG, Escherichia coli L-arabinose transport ATP-binding protein (504 aa), fasta scores: opt: 305 z-score: 368.0 E(): 3.6e-13, 45.8% identity in 96 aa overlap. Similar to many others from S.coelicolor e.g. TR:O50504 (EMBL:AL009199) S.coelicolor probable ABC-transport system ATP binding protein (260 aa) (55.2% identity in 96 aa overlap). Contains pfam match to PF00005 ABC\_tran, ABC transporter and PS00017 ATP/GTP-binding site motif A (P-loop)"  
 /codon\_start=1  
 /transl\_table=11  
 /label="SCF43A.01c"  
 /product="putative ABC transporter ATP-binding protein"  
 /protein\_id="CAB48888.1"  
 /db\_xref="GI:5459215"  
 /translation="MTFSDRGAPVLALRDVSKFGAVRALQVSLFPGVEVHALA GENGAGKSTLIKTLAGVHRPDAGHVLIDGAPVVFHGGPDANDAGIAVIIOETLFPOL SI"  
 1..101  
 /note="overlap with Streptomyces coelicolor cosmid F43 from 22296 to 22396"  
 complement(157..180)  
 /gene="SCF43A.01c"  
 /note="PS00017 ATP/GTP-binding site motif A (P-loop)"  
 577..581  
 /note="possible RBS"  
 589..1749  
 /gene="SCF43A.02"

## CDS

589..1749  
 /gene="SCF43A.02"  
 /note="SCF43A.02, possible sugar isomerase, len: 386 aa; weakly similar to many xylose isomerases (divergent at N- and C-termini) e.g. SW:XYLA\_AMPSP (EMBL:M15050), xyIA, Ampullariaella sp. xylose isomerase (394 aa), fasta scores: opt: 189 z-score: 229.6 E(): 1.8e-05, 26.9% identity in 219 aa overlap"  
 /codon\_start=1  
 /transl\_table=11  
 /label="SCF43A.02"  
 /product="putative sugar isomerase"  
 /protein\_id="CAB48889.1"  
 /db\_xref="GI:5459216"  
 /translation="MTELAAYKAALKQAVETPSWAYGNSGTRFKVFAQGPVPRDPFE KIDDAAVHEFTGAPTVALHPMDRVEDYAAALAAEKGRVIGAINSTFRQDDDYR LGSICHDDAAVRRAKVAHLLCEVDIMDATGSDRLKWFADGTNPGQDDIRSDRLA EGLAETVREGQRMLEYLKLFEPAFTYTDVDMGTAYAHCLKLGEKAQVVYDIGHH APTNIEFIVATLLEKIGLGGDFNSFYADDDLMVGAADPFQLFRIMYVVRGGFT SDVAFMLDQCHNLEAKIPALIRSNVYOETAKALINDGTALAEQAAGDVLENAVLT MDANTDYRCLLREVRESGLDPEFPMKAYRSCGWAEKVVAERIGGQQAGWA"

## gene

1868..3907  
 /gene="SCF43A.03"

## CDS

1868..3907  
 /gene="SCF43A.03"  
 /note="SCF43A.03, probable oxidoreductase, len: 679 aa; similar to many members of the short-chain dehydrogenase/reductase family e.g. SW:YUXG\_BACSU (EMBL:J93938) Bacillus subtilis hypothetical oxidoreductase (689 aa), fasta scores: opt: 1727 z-score: 1840.8 E(): 0, 40.3% identity in 677 aa overlap. C-terminal half is similar to acetoin dehydrogenases e.g. SW:BUDC\_KLEPN (EMBL:D86412), budC, Klebsiella pneumoniae acetoindiacetyl reductase (256 aa) (35.7% identity in 258 aa overlap). C-terminal half is similar to putative dehydrogenases from S.coelicolor e.g. TR:O88068 (EMBL:AL031541) S.coelicolor probable dehydrogenase (260 aa) (34.3% identity in 268 aa overlap). Contains probable coiled-coil from 307 to 334 (28 residues) (Max score: 1.492, probability 0.94). Contains Pfam match to entry PF00106 adh\_short, short chain dehydrogenase"  
 /codon\_start=1  
 /transl\_table=11  
 /label="SCF43A.03"  
 /product="putative oxidoreductase"  
 /protein\_id="CAB48890.1"  
 /db\_xref="GI:5459217"  
 /translation="MGTHPSAALLARSHRIGADPRNTYAGGNASAKGTGTPVTGG EVELMMVKGSGDGLTGTAGLAVLRDMRLKDYVPGVEHEDEMVAADFYLHGKG GAPSIDTFAMHGLVEAAHYDHLHPDGSIGALACADGEKLTAECDTVVWPWRPGF QLGIDIAVKEANPQAVGCVLGHGITAAGDSTDECEKNSLHIITAEATLARGKAE PPGVPIYAAALPAEERRAALAPYVVRGLASRDRAQVGHFTDADVLDLFLARAEHP RLAAAGTSCDPHFTRTKVRPLVDVAPTALEAVARLAKELHAAYREYAAHYERHAE PPSPMRGADPAIVLPVGVGMSEFGCKQTAHVAGFEYVNAINMARGAASVYAPIE ESEKRIETWALEAKLRMRPKPLATRLVLTGAGSGIGAKIARLIVDEGACVVVA DLNAENAAVAEEGGDKXAVAVTVDVTEEQTAAPAAALAFGGVDLVVYNNAGLSI SKPLETSKMDQLQHDIMARGSLVRSREARVMTAQELGGDITVIASKNAVAGPNN IAYSTAKDQAQHVRLAALAEHGEIRVNGPNDGVYRGSGIFAGGWAARAAYGVYP EKLKGEFTAQRTLLKREVLPEHVANAVFALTGGDLTHTTGLHVPVDAGVAAFLR"

## misc\_feature

3128..3685  
 /gene="SCF43A.03"  
 /note="pfam match to entry PF00106 adh\_short, short chain dehydrogenase, score 197.80, E-value 1.7e-55"  
 3904..5358  
 /gene="SCF43A.04"  
 3904..5358  
 /gene="SCF43A.04"

## gene

3904..5358  
 /gene="SCF43A.04"

## CDS

3904..5358  
 /gene="SCF43A.04"  
 /note="SCF43A.04, probable rhamnose kinase, len: 484 aa; similar to rhamnose kinases e.g. SW:RHAB\_ECOLI (EMBL:L19201), rhaB, Escherichia coli rhamnulokinase (489 aa), fasta scores: opt: 1136 z-score: 1280.7 E(): 0, 41.5% identity in 468 aa overlap. Also similar to many other carbohydrate kinases e.g. SW:FUCK\_ECOLI (EMBL:X15025),





```

VERSION      X16144.1  GI:47976
KEYWORDS     acyl carrier protein; ketoacyl reductase; polyketide synthase.
SOURCE       Streptomyces violaceoruber.
ORGANISM     Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
              Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE    1 (bases 1 to 5880)
AUTHORS      Sherman,D.H.
TITLE        Direct Submission
JOURNAL      Submitted (16-AUG-1989) Sherman D.H., John Innes Institute and AFRC
              Institute of Plant Science Research, Colney Lane, Norwich NR4 7UH,
              UK
REFERENCE    2 (bases 1 to 6793)
AUTHORS      Sherman,D.H., Malpartida,F., Bibb,M.J., Kieser,H.M., Bibb,M.J. and
              Hopwood,D.A.
TITLE        Structure and deduced function of the granaticin-producing
              polyketide synthase gene cluster of streptomyces violaceoruber Tu22
JOURNAL      EMBO J. 8 (1989) In press
FEATURES     Location/Qualifiers
              1..6793
               /organism="Streptomyces violaceoruber"
               /strain="Tu22"
               /db_xref="taxon:1935"
               /clone_lib="EMBL4 genomic"
               complement(78..827)
               /note="ketoacyl reductase (AA 1-289)"
               /codon_start=1
               /transl_table=11
               /protein_id="CAA34262.1"
               /db_xref="GI:47977"
               /db_xref="SWISS-PROT:P16543"
               /translation="NATDAPEAPVALVTGGSSGIGQTVAFORLAEGYRVVYNARSVE
               DGEKTAALPDALYRADYSEADARRLYDRAVEHYGRDLVYNNAGRTRAIRAPADLA
               AAPTEWREILGNVIGTQTVAAMPHLARGSGSVYNNVSSIAGSPAGSSIPYAVS
               NGHGATNRLANTVGPVAVNAVAPGLIETPTWQNSDFFAPTIAEHVROTPTPLRTGR
               PEDVAEAVLGLVRATYTTGOVLLVDGGAHL"
               complement(827..1645)
               /note="ketoacyl reductase (AA 1-273)"
               /codon_start=1
               /transl_table=11
               /protein_id="CAA34263.1"
               /db_xref="GI:47978"
               /db_xref="SWISS-PROT:P16542"
               /translation="MTATATATATPGTAAPKPAVALVTGATSGIGLAIARLAAALGART
               FLCARDEERLAQTVKELRGEGFDVGTCDVADPAOIRAVVAAVORYGVVDILVNN
               GRSGGATAEIADLWLDVITNTLSVFLMTKEVNLAGMLAKKRGRIINATSGKQ
               GYVHAVPSASKHGVYGLTKALGLELARTGITVNAVCPGFVETPMARVREHAGINQ
               VSEETFDRTITNRVPLGRVETREVAAMVEYLVAADAAVAATQAALNVCGLIGN"
               1914..3179
               /note="polyketide synthase ketoacyl synthase (AA 1-421)"
               /codon_start=1
               /transl_table=11
               /protein_id="CAA34264.1"
               /db_xref="GI:581781"
               /db_xref="SWISS-PROT:P16540"
               /translation="MTRRVITGVGVAPGGSGTKEFWDLTLTAGRTATRPISFFDASP
               FRSRAGIDFDVAEGSPREVRMDRATOFACVACTDADLSDGLTGDALDPSRIGV
               ALGSVATSENLVYMSDREWLVDPAHLPMFMDYALSPGVPMAEVAWAAGAG
               PVTWSDGCTGLDSVGVAGTREGSADVAGAADTPVSPVIVACFDAIKATTPRN
               DDPNARSREFDTRNGFVLAEGAAFMVLEEYEAQRGAHIYAEVGYATRSQAYHMT
               GLKDGRENAESIRALDELDTADVYNVNAHSGTKQNDRHETAFKSLGEHAY
               VPVSSIKMGSHSLGASIEIAASVLAIEHNVPPTANLHTPDPCDLDLYVPLTARE
               QRVDVLTVTGSGFGGFSQANVLHRPEAA"
               3176..4423
               /note="put. ketoacyl reductase (AA 1-414)"
               /codon_start=1
               /transl_table=11
               /protein_id="CAA34265.1"
               /db_xref="GI:581782"
               /db_xref="SWISS-PROT:P16541"
               /translation="MSTPDRRAVVTGLSVAAPGGLGTERYWKSLLTGENGIAELSRF
               DASRYPRLAGQIDDFEASEHLPSRLLPQTQDSTRYALAAADWALADAGVPESGLDD
               YDLGVVTSTAAGGDFDTHREFHKLWSQGPAYVSVYVESPAFYAVNTGTGISRNTMRGP

```

```

SAALVGEQAGGLDAIGHARTVRRGFCWCSAVASTRRSTRGASSSOLSGGLYSTVADP
ERAVLPEDVASGVYPCGEGAVLIVEDADSARARGAERTYVRSPLRRDPAPGSGRPPA
LGRAELALAEAGLTADISVFDAGAGVPELDRAEADTLARLFGPRGVPVTPAKALT
GRLCAGGSPADLAALLALROQVIPATGRHRAVPDAYALDLYTIGRPREAALSALVLA
RGRGFNSAVVTLRSGDHRPPI"
4505..4762
 /note="acyl carrier protein (AA 1-85)"
 /codon_start=1
 /transl_table=11
 /protein_id="CAA34266.1"
 /db_xref="GI:47981"
 /db_xref="SWISS-PROT:P12885"
 /translation="MARLTLDGKRLILVACAGEDDGVLDLSDGILDITFEEELGYDSLAL
 MESASRIEREELGVALADGDINEELTPRVLLDLVNGAQAEAA"
4882..5856
 /note="URF (AA 1-324)"
 /codon_start=1
 /transl_table=11
 /protein_id="CAA34267.1"
 /db_xref="GI:47982"
 /db_xref="SWISS-PROT:P16560"
 /translation="MVQPAATPVSLPSTVHRSEHTVTVAAPPEALYALVADVTRWPA
 VFETVVRHLAREGRTFEIWAENGEIAHWSRRVLDVPRVYVFFQESRPPVT
 SMGGMFLRPLADGRTEIVLRHRTVADDDPAAVARTEELADNRNARELGALAALET
 GHVDELVFSFTDPLQGAARDAYTEVERAEWAELPHVAOCGRAGTGLEQWLE
 MDVTADGSHHTTSRARCIPAPENIAYNEQRTPLYSHGSGSEWTFPAQTPEGPVATARH
 TVAVDPGITEVLPDADLADARLADRLGRNSLATLRHAAEAORA"
BASE COUNT 903 a 2666 c 2352 g 872 t
ORIGIN

alignment_scores:
      Quality: 103.50      Length: 69
      Ratio: 2.797      Gaps: 3
      Percent Similarity: 53.623      Percent Identity: 37.681

alignment_block:
US-09-135-238B-2_COPY_273_390 x SVGRA/rev ..
Align seg 1/1 to reverse of: SVGRA from: 1 to: 6793
23 AlaLeuGluSerSerGlnArgProArgGlySerProArgProArgSerG1 39
||| :||:||||| :||| :||:|||||
1870 GCCCGCGCTACCTCACTGACACCCCGGCACACTCCCGCGGCC.....1829
39 nAsuAsnIleTy:SerAlaCy:ProArgAlaArgGlyAlaAspAlaA 56
||| ||| ||| ||| ||| ||| ||| ||| |||
1828 .....TGCTCGAGAGCGGCTCGCGCGCGGATGGGCG 1798
56 laGlyThrGlyGluAlaProValProGlyProGlyAlaProLeuPro... 71
||| ||| :||: ||| ||| :||:|||||
1797 GCGGAGAGGCGAGTGGTGGTCTCGTCCCGCGGCGAACGCCCATCCCGCTC 1748
72 .....ProAlaProLeuGlnValSerGluSerProTrpLeuHi 84
||| ||| ||| ||| ||| ||| ||| ||| |||
1747 TCCATCGGGCTTCACGCCCGCGCTC.....GAGGCCCGCTTGGCGGACG 1707
84 saLaPro 86
:|:|:|:|
1706 CTCTCCC 1700
seq_name: gb_ro:BTFRWNP2
seq_documentation_block:
LOCUS      BTFRWNP2 13187 bp DNA ROD 11-DEC-1996
DEFINITION R.norvegicus PRM1, PRM2, PRM3 and TNP2 genes.
ACCESSION 246939
VERSION 246939.1 GI:1359527
KEYWORDS PRM1 gene; PRM2 gene; PRM3 gene; protamine; TNP2 gene; transition
              protein 2.
SOURCE      Rattus norvegicus
ORGANISM    Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

```





JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Unpublished  
2 (bases 1 to 37200)  
Bentley, S.D., Parkhill, J., Barrell, B.G. and Rajandream, M.A.  
Direct Submission  
Submitted (17-MAR-1999) Streptomyces coelicolor sequencing project,  
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge  
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.  
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,  
Colney, Norwich, Norfolk NR4 7UH, UK  
3 (bases 1 to 37200)

REFERENCE  
AUTHORS

Redenbach, M., Kieser, H.M., Denapate, D., Eichner, A., Cullum, J.,  
Kinashi, H. and Hopwood, D.A.  
A set of ordered cosmids and a detailed genetic and physical map  
for the 8 Mb Streptomyces coelicolor A3(2) chromosome  
Mol. Microbiol. 21 (1), 77-96 (1996)  
97000351

TITLE

JOURNAL  
MEDLINE  
COMMENT

On Mar 24, 1999 this sequence version replaced gi:4481931.  
Notes:  
Streptomyces coelicolor sequencing at The Sanger Centre is funded  
by the BBSRC.

Details of S. coelicolor sequencing at the Sanger Centre are  
available on the World Wide Web.  
(URL: [http://www.sanger.ac.uk/Projects/S\\_coelicolor/](http://www.sanger.ac.uk/Projects/S_coelicolor/)) CDS are  
numbered using the following system eg SC7B7.01c. SC (S.  
coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary  
strand).

The more significant matches with motifs in the PROSITE database  
are also included but some of these may be fortuitous. The length  
in codons is given for each CDS.

Usually the highest scoring match found by fasta -o is given for  
CDS which show significant similarity to other CDS in the database.  
The position of possible ribosome binding site sequences are given  
where these have been used to deduce the initiation codon. Gene  
prediction is based on positional base preference in codons using a  
specially developed Hidden Markov Model (Krogh et al., Nucleic  
Acids Research, 22(22):4768-4778(1994)) and the FramePlot program  
of Bibb et al., Gene 30:157-66(1984) as implemented at  
<http://www.nih.gov.jp/>

<http://cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the  
correct initiation codon. Where possible we choose an initiation  
codon (atg, gtg, ttg or att) which is preceded by an upstream  
ribosome binding site sequence (optimally 5-13bp before the  
initiation codon). If this cannot be identified we choose the most  
upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the  
sequenced clone. It may be shorter because we only sequence  
overlapping sections once, or longer, because we arrange for a  
small overlap between neighbouring submissions. Cosmid E63 lies  
between E8 and E29 on the AseI-E genomic restriction fragment.

FEATURES  
source

1. .37200  
/organism="Streptomyces coelicolor A3(2)"  
/strain="A3(2)"  
/db\_xref="taxon:100226"  
/clone="cosmid E63"  
complement(1. .138)  
/gene="SCE63.06"  
complement(<1. .138)  
/gene="SCE63.06"

/note="SCE63.06, partial CDS, possible aminotransferase,  
len: >45aa; similar to the N-terminal region of TR:052815  
(EMBL:AJ223998) a protein similar to aminotransferase from  
a cluster of genes involved in the biosynthesis of a  
vancomycin group antibiotic in Amycolatopsis orientalis  
(Actinomycete) (438 aa) fasta scores; opt: 87, z-score:  
147.0. E1: 0.69, (45.2% identity in 31 aa overlap). The  
remainder of this CDS lies on the neighbouring  
Streptomyces coelicolor cosmid E8."

/codon\_start=1  
/transl\_table=11  
/label="SCE63.06"

/product="putative aminotransferase"  
/protein\_id="CAB38521.1"

gene

CDS

/db\_xref="GI:4481937"  
/translation="MTTITPARDRGVLARTALHPLSLAPLLDTMNFLEVTLYRYPQ  
AI"  
complement(135. .1268)  
/gene="SCE63.05" .1268)  
complement(135. .1268)  
/gene="SCE63.05" .1268)  
/note="SCE63.05, probable glycolate oxidase, len: 377aa;  
similar to many both prokaryotic and eukaryotic egs.  
TR:052792 (EMBL:AJ223998) similar to glycolate oxidase  
from a cluster of genes involved in the biosynthesis of a  
vancomycin group antibiotic in Amycolatopsis orientalis  
(Actinomycete) (357 aa) fasta scores; opt: 995, z-score:  
998.4. E1: 0, (47.5% identity in 343 aa overlap) and  
SW:GOX\_SPIOI glycolate oxidase from Spinacia oleracea  
(spinach) (369 aa) fasta scores; opt: 917, z-score: 920.7,  
E1: 0, (41.9% identity in 360 aa overlap). Contains Pfam  
match to entry PF01070 FMN\_dh, FMN-dependent de-  
hydrogenase, score 412.60, E-value 3.6e-120 and P500557  
FMN-dependent alpha-hydroxy acid dehydrogenases active  
site."  
/codon\_start=1  
/transl\_table=11  
/label="SCE63.05"  
/product="putative glycolate oxidase"  
/protein\_id="CAB38520.1"  
/db\_xref="GI:4481936"  
/translation="MREPLTLDLDFARLARGQLPAATWDFIAGGAGRETLAANEAVFG  
AVLRRLPALGIEEDPTSEVIGSRNPAPVGIAPVAVYGLAHPDGPATAAAGALGL  
PLVYSTFAGRSLEFVARAASAPILQLYCFRDHETTLGLARRADSGYQALVLTDP  
FTGRLRLDNLNGFAVPIHTPANTLTGAAGSATPGAHSLAFDRRLDSEFVARLGRA  
SGELVLAKGYLTADPAEAAVAGVIVSNHGQQLDGPATLEALPLEVVSVAVRGHC  
PVLIDGGVGTGADVLAALGARGAVLVRGALYALVGGAGVRRMLTLITDFADTM  
VLTHAAATGRTIGPDTLAPPHAPPHGPTAPRPAHRDRSHG"  
complement(213. .1229)  
/gene="SCE63.05"  
/note="Pfam match to entry PF01070 FMN\_dh, FMN-dependent  
de hydrogenase, score 412.60, E-value 3.6e-120"  
complement(507. .527)  
/gene="SCE63.05"  
/note="P500557 FMN-dependent alpha-hydroxy acid  
dehydrogenase ses active site."  
/label="\*"  
complement(1418. .2533)  
/gene="SCE63.04"  
complement(1418. .2533)  
/gene="SCE63.04"  
/note="SCE63.04, probable 4-hydroxyphenylpyruvic acid  
dioxxygenase, len: 371aa; similar to many both prokaryotic  
and eukaryotic egs. TR:052791 (EMBL:AJ223998) similar to  
hydroxyphenyl pyruvate dioxxygenase from a cluster of genes  
involved in the biosynthesis of a vancomycin group  
antibiotic in Amycolatopsis orientalis (Actinomycete) (357  
aa) fasta scores; opt: 989, z-score: 1130.3. E1: 0,  
(48.7% identity in 355 aa overlap) and SW:HPD\_MOUSE hpd,  
4-hydroxyphenylpyruvic acid dioxxygenase from Mus musculus  
(mouse) (392 aa) fasta scores; opt: 610, z-score: 698.5,  
E1: 1.3e-31, (31.6% identity in 361 aa overlap)."  
/codon\_start=1  
/transl\_table=11  
/label="SCE63.04"  
/product="Putative 4-hydroxyphenylpyruvic acid  
dioxxygenase"  
/protein\_id="CAB38519.1"  
/db\_xref="GI:4481935"  
/translation="MLPPFPFLHWRAMPDIAVAVLVADREASGFLVDSLGFPV  
LAVAGPATGTHDRRSTVRSGEVTLVVTQALADPTVARYVEHGDSIADLAFGCDV  
RSCFDRAVLAGAENLAQPTPSHRAGQDAFWATVSGGDIRHTILVPAADSGAGLLPPD  
RDWALLPATGRTGPRPLDHWAVACSGTSTRAEFYAEFADMPYSEIYEVGEQA  
MDSIVVRNAGGGITFLIEPDDTRVFGQIDQFLSHADGPGVHLAFVDDIYGVSRSL  
GDRGVAFLTPGAYDILLTFRVGAMADAIEDLRETNVLADRDDEWGYLLQIFTRSPYR  
GTLFYETIQNRNARGFGSSNIKALYEAVEREREVAGR"  
2802. .25193  
gene  
misc\_feature  
misc\_feature  
gene  
CDS  
gene  
CDS  
gene  
CDS  
gene











OM of: US-09-135-238B-2\_COPY\_18\_253 to: N\_Geneseq\_36.\* out\_format : pfs

Date: Sep 12, 2000 6:55 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODEL=framed.p2n.model -DEV=xlp
-Q/cgn2_1/USPTO.spool/US09135238/runat_24082000_091324_17918/app_query.fasta_1.1282
-DB=N_Geneseq_36 -QPMT=fastap -SUFFIX=p2n.rng -GAPOP=12.000
-GAEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAEXT=0.050 -XGAPOP=10.000 -XGAEXT=0.500
-FGAPOP=6.000 -FGAEXT=7.000 -YGAPOP=10.000 -YGAEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blonsum62
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=100000
-USER=US09135238 @CGN1_1_123 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPVY
-WAIT -THREADS=1
```

Search information block:

```
Query: US-09-135-238B-2_COPY_18_253
Query length: 236
Database: N_Geneseq_36.*
Database sequences: 311585
Database length: 125096042
Search time (sec): 121.040000
```

score\_list:

Sequence	Strd Orig	ZScore	EScore Len	Documentation
N_Geneseq_36:X28178	+ 1254.00	1961.26	1.2e-101	2040 ! Human PIGRL-1 coding sequence.
N_Geneseq_36:X28179	+ 1029.50	1613.23	2.9e-82	1047 ! Human PIGRL-1 coding sequence.
N_Geneseq_36:X13290	+ 197.00	287.40	2.1e-08	3095 ! Mouse poly-immunoglobulin rece
N_Geneseq_36:X15422	+ 188.00	250.87	2.3e-06	29392 ! Mouse poly Ig receptor protei
N_Geneseq_36:X31291	+ 185.50	268.69	2.3e-07	3269 ! Rat poly-immunoglobulin recept
N_Geneseq_36:X07407	+ 180.50	266.49	3.0e-07	1839 ! Human secretory immunoglobulin
N_Geneseq_36:X30857	+ 180.50	285.99	3.2e-07	1936 ! Secretory component DNA fragme
N_Geneseq_36:X30858	+ 180.50	265.51	3.5e-07	2031 ! Secretory component DNA fragme
N_Geneseq_36:X31287	+ 180.50	261.92	5.5e-07	2919 ! Human poly-immunoglobulin rece
N_Geneseq_36:X31288	+ 179.50	258.18	8.8e-07	3630 ! Bovine poly-immunoglobulin rec
N_Geneseq_36:X31289	+ 175.50	238.22	8.8e-07	1911 ! cDNA for human immunity relate
N_Geneseq_36:X14432	+ 166.50	244.18	5.3e-06	1876 ! Clone 1 for truncated poly Ig-
N_Geneseq_36:X14439	+ 166.50	237.95	1.2e-05	3517 ! Rabbit poly-immunoglobulin rec
N_Geneseq_36:X14499	+ 145.00	210.22	0.0004	1876 ! Clone 2 for truncated poly Ig-
N_Geneseq_36:X28203	+ 119.50	174.90	0.0385	1137 ! Human kidney injury related mo
N_Geneseq_36:X96035	+ 117.50	167.21	0.1031	1795 ! Human kidney injury related mo
N_Geneseq_36:X87689	+ 116.00	176.30	0.0321	565 ! EST clone EN260. New polynucle
N_Geneseq_36:X96033	+ 111.50	156.26	0.4203	2084 ! Rat kidney injury related mole
N_Geneseq_36:X96032	+ 111.50	154.19	0.5476	2566 ! Rat kidney injury related mole
N_Geneseq_36:X34234	+ 107.50	156.70	0.3969	1053 ! Human secreted protein gene 16
N_Geneseq_36:X34235	+ 107.50	156.50	0.4075	1075 ! Human secreted protein gene. N
N_Geneseq_36:X34169	+ 106.50	154.65	0.5168	1105 ! Human secreted protein gene 16
N_Geneseq_36:X18694	+ 104.50	131.05	10.66	8690 ! Plasmid PD17-hum14.H1. Inhibit
N_Geneseq_36:X41432	+ 104.00	148.16	1.19	1427 ! Plasmid Hu19Hcpcd encoding an
N_Geneseq_36:X41429	+ 103.00	146.58	1.45	1427 ! Plasmid Hu19Hcpcd encoding an
N_Geneseq_36:X41431	+ 103.00	146.58	1.45	1427 ! Plasmid Hu19Hcpcd encoding an
N_Geneseq_36:X18598	+ 102.50	128.32	15.13	8321 ! Plasmid expressing hBR95-A. In
N_Geneseq_36:X15932	+ 102.00	129.89	12.37	6557 ! Anti-IgE VH expression vector.
N_Geneseq_36:X38716	+ 101.50	136.24	5.48	3190 ! CD22-beta gene. Inhibition of
N_Geneseq_36:X62510	+ 101.00	143.40	2.19	1431 ! Primatized anti-human B7.1 and
N_Geneseq_36:X35485	+ 101.00	143.40	2.19	1431 ! Macaque primatized 7C10 heavy
N_Geneseq_36:X03509	+ 100.50	141.39	2.83	1617 ! Antibody D heavy chain. Prodn.
N_Geneseq_36:X41427	+ 100.00	127.15	17.57	6284 ! Plasmid Hu19AHcpcd encoding an
N_Geneseq_36:X60665	+ 99.50	130.46	11.50	4157 ! Human CD100 antigen cDNA. Nucl
N_Geneseq_36:X24074	+ 99.00	144.68	1.86	914 ! Human EST R2810 DNA. Complexes
N_Geneseq_36:X66631	+ 99.00	137.59	4.61	1869 ! Human immunoglobulin heavy cha
N_Geneseq_36:X80470	+ 98.50	145.41	1.69	784 ! Anti-phenylloxazolone antibody c
N_Geneseq_36:X80580	+ 98.50	142.42	2.48	1060 ! Kidney injury associated molec
N_Geneseq_36:X07474	+ 98.50	125.03	23.06	6127 ! Mus musculus anti-IgE antibody
N_Geneseq_36:X049834	+ 97.00	137.39	4.72	1386 ! Anti-HIV-1 recombinant antibod

N\_Geneseq\_36:Q12632 + 97.00 132.10 9.31 2364 ! CDA-specific CDR-grafted he  
N\_Geneseq\_36:Q49944 + 96.50 135.33 6.15 1576 ! Human anti-HBs heavy chain.  
N\_Geneseq\_36:X52228 + 96.50 131.63 9.90 2290 ! Protein PRO27 cDNA clone D  
N\_Geneseq\_36:V38929 + 96.00 132.49 8.86 1938 ! A33 chimeric receptor DNA.  
N\_Geneseq\_36:N81713 + 96.00 127.69 16.40 3146 ! Clone contg. thrc gene. Thr

seq\_name: N\_Geneseq\_36:X28178

seq\_documentation\_block:

ID X28178 standard; cDNA; 2040 BP.  
AC X28178;  
DT 16-JUN-1999 (first entry)  
DE Human PIGRL-1 coding sequence.  
KW PIGRL-1; human; autoimmune disease; hyper-IgM Immunodeficiency; HIM;  
KW X-linked Severe Combined Immunodeficiency; XSCID; IgA deficiency;  
KW diagnosis; therapy; ss.  
OS Homo sapiens.  
PN EP-905238-A2.  
PD 31-MAR-1999.  
PE 14-AUG-1998; 306487.  
PR 30-OCT-1997; US-961564.  
PR 25-AUG-1997; US-056935.  
PA (SMIK ) SMITHKLINE BEECHAM CORP.  
PI Sweet RW, Truneh A, Wu S;  
DR WPI: 99-192666/17.  
DR P-PSDB: Y05001.  
PT New polypeptides encoding PIGRL-1 useful for treating diseases such  
as X-linked Severe Combined Immunodeficiency  
PS Claim 4; Page 18-19; 26pp; English.  
CC This sequence encodes the human PIGRL-1 protein of the invention.  
CC Autoimmune diseases involving altered expression or activity of PIGRL-1  
CC may include Hyper-IgM Immunodeficiency (HIM), X-linked Severe Combined  
CC Immunodeficiency (XSCID) and IgA deficiency. These diseases can be  
CC diagnosed or susceptibility to them predicted by: (1) determining whether  
CC there is a mutation in the genomic copy of the gene encoding PIGRL-1; or  
CC (2) measuring the amount of PIGRL-1 in a sample derived from the patient.  
CC Patients deficient in PIGRL-1 can be treated by administering either the  
CC PIGRL-1 DNA or its complement or an agonist of PIGRL-1 to the patient.  
CC Patients with excessive expression or activity of PIGRL-1 can be treated  
CC by administering an antagonist of PIGRL-1, an antisense nucleic acid  
CC molecule which inhibits the expression of PIGRL-1 or administering  
CC sufficient PIGRL-1 to compete with the endogenous activity. PIGRL-1 can  
CC be used to identify its agonists by contacting a cell expressing PIGRL-1  
CC with a candidate compound in the presence of a signal system and noting  
CC the candidate as an agonist if a signal is produced. The same method can  
CC be used to identify antagonists of PIGRL-1 but the presence of an  
CC antagonist is indicated by a decrease in production of the signal.  
CC Antibodies against PIGRL-1 may be used to isolate or identify clones  
CC expressing PIGRL-1. Polynucleotides encoding PIGRL-1 may be used to  
CC identify chromosomal mutations in the gene encoding PIGRL-1 in patients.  
CC This information may then be correlated with the incidence of autoimmune  
CC disease in those patients to identify whether the mutation causes the

alignment\_scores:

Quality: 1254.00 Length: 236  
Ratio: 5.314 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-135-238B-2\_COPY\_18\_253 x X28178 ..  
Align seg 1/1 to: X28178 from: 1 to: 2040  
1 Arg1LeuProGluValLysValGluGlyLeuGlySerValth 17  
|||||  
197 AGGATCTCCAGAGTAAGGTAGAGGGAGCTGGCGGATCATGTTAC 246  
17 r1LeuLysCysProLeuProGluMethHisValArg1LeuTyrLeuCysArg 34  
|||||  
247 CATCAAGTGGCCACTTCTCTCAATGCATGTGAGGATATATCTCTGCGCGG 296

34 luMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsnPhe 50  
 |||||  
 297 AGATGGCTGGATCTGGAACATGTGTACCGTGTATCCACCACTTC 346  
 51 IleLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProArgLy 67  
 |||||  
 347 ATCAAGGCAGAAATACAAGGGCCGAGTTACTCTGAAGCAATATCCACGCAA 396  
 67 sAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerGlyV 84  
 |||||  
 397 GAATCTGTCTCTAGTGGAGGTAAACACAGCTGACAGAAAGTGACAGCGGAG 446  
 84 alTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLys 100  
 |||||  
 447 TCTATGCTCGGAGCGGGCATGAACACAGACCGGGGAAAGACCCAGAAA 496  
 101 ValThrLeuAsnValHisSerGluTyrGluProSerTrpGluGlnPr 117  
 |||||  
 497 GTCAACCTGAATGTCCACAGTGAATACGAGCCATCATGGGAGAGCACC 546  
 117 oMetProGluThrProLysTrpPheHisLeuProTyrLeuPheGlnMetP 134  
 |||||  
 547 AATGCTGAGACTCCAAAATGGTTTCACTGCGCCTATTGTTCAGATGC 596  
 134 roAlaTyrAlaSerSerSerLysPheValThrArgValThrThrProAla 150  
 |||||  
 597 CTGCATATGCCAGTCTCTCCAAATTCGTAAACAGAGTTACACACACAGCT 646  
 151 GlnArgGlyLysValProProValHisHisSerSerProThrThrGlnIl 167  
 |||||  
 647 CAAGGGCAGAGTCCCTCCAGTTCACACTCTCTCCCCACACCCCAAT 696  
 167 eThrHisArgProArgValSerArgAlaSerSerValAlaGlyAspLysP 184  
 |||||  
 697 CACCACCCGCTCGAGTGTCCAGAGCATCTTCAGTAGCAGGTGACAAAG 746  
 184 roArgThrPheLeuProSerThrThrAlaSerLysLysSerAlaLeuGlu 200  
 |||||  
 747 CCCGAACCTCTCTGCCATCCTACACTACAGCCTCAAAAATCTCAGCTCTGGAG 796  
 201 GlyLeuLeuLysProGlnThrProSerTyrAsnHisHisThrArgLeuHi 217  
 |||||  
 797 GGGCTGCTCAAGCCCGCAGAGCCGACGTACACACACACACAGGCTGCA 846  
 217 sArgGlnArgAlaLeuAspTyrGlySerGlnSerGlyArgGluGlyGlnG 234  
 |||||  
 847 CAGGCAGAGACACTGGACTATGGCTCACAGTCTGGGAGGGAAGGCCAAG 896  
 234 lyPheHis 236  
 |||||  
 897 GATTCAC 904

seq\_name: N\_Geneseg\_36:X28179  
 seq\_documentation\_block:  
 ID X28179 standard; cDNA; 1047 BP.  
 AC X28179;  
 DT 16-JUN-1999 (first entry)  
 DE Human PIGRL-1 coding sequence fragment.  
 KW PIGRL-1; human; autoimmune disease; hyper-IgM Immunodeficiency; HIM;  
 KW X-linked Severe Combined Immunodeficiency; XSCID; IgA deficiency;  
 KW diagnosis; therapy; ss.  
 OS Homo sapiens.  
 PN EP-905236-A2.  
 PD 31-MAR-1999.  
 PF 14-AUG-1998; 306487.  
 PR 30-OCT-1997; US-961564.  
 PR 25-AUG-1997; US-056935.  
 PA (SMK) SMITHKLINE BEECHAM CORP.  
 PI Sweet RW, Truneh A, Wu S;  
 DR WPI; 99-192666/17.  
 DR P-PSDB; Y05002.

PT New polypeptides encoding PIGRL-1 useful for treating diseases such  
 as X-linked Severe Combined Immunodeficiency  
 PS Disclosure: Page 8; 26pp; English.  
 CC This sequence encodes the human PIGRL-1 protein of the invention.  
 CC Autoimmune diseases involving altered expression or activity of PIGRL-1  
 may include hyper-IgM immunodeficiency (HIM), X-linked Severe Combined  
 CC Immunodeficiency (XSCID) and IgA deficiency. These diseases can be  
 CC diagnosed or susceptibility to them predicted by: (1) determining whether  
 CC there is a mutation in the genomic copy of the gene encoding PIGRL-1; or  
 CC (2) measuring the amount of PIGRL-1 in a sample derived from the patient.  
 CC Patients deficient in PIGRL-1 can be treated by administering either the  
 CC PIGRL-1 DNA or its complement or an agonist of PIGRL-1 to the patient.  
 CC Patients with excessive expression or activity of PIGRL-1 can be treated  
 CC by administering an antagonist of PIGRL-1, an antisense nucleic acid  
 CC molecule which inhibits the expression of PIGRL-1 or administering  
 CC sufficient PIGRL-1 to compete with the endogenous activity. PIGRL-1 can  
 CC be used to identify its agonists by contacting a cell expressing PIGRL-1  
 CC with a candidate compound in the presence of a signal system and noting  
 CC the candidate as an agonist if a signal is produced. The same method can  
 CC be used to identify antagonists of PIGRL-1 but the presence of an  
 CC antagonist is indicated by a decrease in production of the signal.  
 CC Antibodies against PIGRL-1 may be used to isolate or identify clones  
 CC expressing PIGRL-1. Polynucleotides encoding PIGRL-1 may be used to  
 CC identify chromosomal mutations in the gene encoding PIGRL-1 in patients.  
 CC This information may then be correlated with the incidence of autoimmune  
 CC disease in those patients to identify whether the mutation causes the  
 CC disease.  
 SQ Sequence 1047 BP; 255 A; 281 C; 276 G; 223 T;

alignment\_scores:  
 Quality: 1029.50 Length: 245  
 Ratio: 4.457 Gaps: 10  
 Percent Similarity: 94.286 Percent Identity: 92.245  
 alignment\_block:  
 US-09-135-238B-2\_COPY\_18\_253 x X28179 ..  
 Align seg 1/1 to: X28179 from: 1 to: 1047  
 1 ArgIleLeuProGluValLysValGluGlyGluLeuGlyGlySerValTh 17  
 |||||  
 201 AGGATCTCCAGAGTAAGTAAGGTAGAGGGGAGCTGGCGGATCAGTTAC 250  
 17 rIleLysCysProLeuProGluMetHisValArgIleTyrLeuCysArgG 34  
 |||||  
 251 CATCAAGTCCCACTTCTCGAAATGTCATGTGAGGATATATCTGTGCGGG 300  
 34 luMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsnPhe 50  
 |||||  
 301 AGATGGCTGGATCTGGAACATGTGTACCGTGTATCCACCACTTC 350  
 51 IleLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProArgLy 67  
 |||||  
 351 ATCAAGGCAGAAATACAAGGGCCGAGTTACTCTGAAGCAATATCCACGCAA 400  
 67 sAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerGlyV 84  
 |||||  
 401 GAATCTGTCTCTAGTGGAGGTAAACACAGCTGACAGAAAGTGACAGCGGAG 450  
 84 alTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLys 100  
 |||||  
 451 TCTATGCTCGGAGCGGAGTGAAGCAATATCCACGAAAGACCCAGAAA 499  
 101 ValThrLeuAsnValHisSerGluTyrGluProSerTrpGluGlnPr 117  
 |||||  
 500 GTCAACCTGAATGTCCACAGTGAATACGAGCCATCATGGGAGAGCACC 549  
 117 oMetProGluThrProLysTrpPheHisLeuProTyrLeuPheGlnMetP 134  
 |||||  
 550 AATGCTGAGACTCCAAAATGGTTTCACTGCGCCTATTGTTCAGATGC 599  
 134 roAlaTyrAlaSerSerSerLysPheValThr..ArgValThrThrProA 150



```

|||||.....|
600 CTGCATATGCGGTCTCTCCACATTCGTAACCCGAGATTACCCACCAG 649
150 la.GlnArgGlyLysValProProValHisHisSerSerProThrThrG1 166
|||.....|
650 CTTCAAGGGGGCAAGGTCCCTCCAGTTCACCACTCTCCCCACCACCCA 699
166 nile.ThrHisArgPro.ArgValSerArgAlaSerSerValAlaGlyAs 182
|||||.....|
700 AATTCACCCACCCGCTTCGAGTGTNCAGAGCATCTTCAGTACAGGTGA 749
182 pLysProArg.ThrPheLeuProSerThrThrAlaSerLysLysSerAla 198
|||||.....|
750 CAAGCCCGCAACTTCTCCCTCCACTACAGCTCAAGCTCAAAAATCTCAGCT 799
199 LeuGluGlyLeuLys.....ProGlnThrProSerTyrAsnHisHisTh 214
|||||.....|
800 CTGAAGGGGTCTTCAAGCCCAAGAGGCCAGCTACAAACAN.CACAC 848
214 rArgLeuHisArgGlnArgAlaLeuAsp...Tyr.GlySerGlnSerGly 229
|||||.....|
849 CAGGCTGCACAGCAGACGACTGGATCTTATGGNTCAGACTGGG 898
230 ArgGluGly.....GlnGlyPhe 235
|||||.....|
899 ...GAGGGGAANGNCCAAGGATTT 919

```

seq\_name: N\_Geneseq\_36:T31290

seq\_documentation\_block:

ID T31290 standard; cDNA; 3095 BP.

AC T31290;

DE Mouse poly-immunoglobulin receptor, cDNA.

KW Mouse; immunoglobulin; receptor; protection protein; mutants;

KW heavy chain; antigen binding domain; protection; pathogen;

KW mucosal; environment; gastrointestinal; passive; immunisation;

KW Guy's 13 antibody; prevention; dental caries; Streptococcus;

KW poly; sorbinus; murine; ss.

OS Mus musculus.

FH Key Location/Qualifiers

FT cds 85..2400

FT /\*tag= a

PN WO9621012-A1.

PD 11-JUL-1996.

PF 27-DEC-1995; U16889.

PR 30-DEC-1994; US-367395.

PR 04-MAY-1995; US-434000.

PA (PLAN-) PLANT BIOTECHNOLOGY INC.

PA (UNKE-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.

PA (PLAN-) PLANET BIOTECHNOLOGY INC.

PI Hiatt AC, Lehner T, Ma JKC;

DR WPI; 96-333987/33.

DR P-PSDB; W03180.

PT Immunoglobulin and protection protein complex and its prodn. in

PT plants - useful for passive immunisation against mucosal antigens,

PT esp. against S. mutans and S. sorbinus to prevent dental caries

PS Disclosure; pages 117-121; 152pp; English.

CC The present sequence encodes the mouse poly-immunoglobulin (Ig)

CC receptor, a portion of which corresp. to residues 1-627, pref.

CC 1-606, or esp. residues 13-45, 1-120, 110-230, 210-340, 320-450,

CC 440-550, 550-606 or 550-627 comprises a protection protein (PP).

CC The Ig of the invention comprises a PP as above in association with

CC an Ig derived heavy chain, having at least a portion of an antigen

CC (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.

CC gastrointestinal, environments, therefore enhancing its

CC effectiveness in passively immunising animals against mucosal

CC pathogens. The Ag binding domain is specifically derived from the

CC Guy's 13 antibody, and the Ig can be used to prevent dental caries

CC by binding, e.g. Streptococcus mutans serotypes c, e and f, or

CC S. sorbinus serotypes d and g.

SQ Sequence 3095 BP; 861 A; 796 C; 784 G; 654 T;

alignment\_scores:

Quality: 197.00 Length: 148

Ratio: 2.165 Gaps: 5

Percent Similarity: 61.486 Percent Identity: 32.432

alignment\_block:

US-09-135-238B-2\_COPY\_18\_253 x T31290 ..

Align seg 1/1 to: T31290 from: 1 to: 3095

```

13 GlyGlySerValThrIleLysCysProLeuProGlu.....Me 25
|||.....|
181 GCGGACTCTGTTTCCATCAGCTGCTACTACCAGACACCTCTGTCAACCG 230
25 ThisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 42
|||.....|
231 GCACACCCGGAATACTGGTCCGACCAAGGAGCC....AGCGGCATGTGCA 277
42 lyThrValValSerThrThrAsnPheIleLysAlaGluTyrLysGlyArg 58
|||.....|
278 CAACGCTCATCTCTCAATGGCTACCTCTCCAAGGAGTATTCAGGCAGA 327
59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75
|||.....|
328 GCCAACCTCATCACTCCAGAGAACACACATTTGTGATTACATTGA 377
75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMetA 92
|||.....|
378 GCAGCTCACCAGGAGCAGACTGGGAGCTACAAAGTGTGGCTGGGT...A 424
92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 108
|||.....|
425 CCAGTAACCCAGGCGCTCTCTTCGATCTCAGCCTGGAGGTC..... 465
109 TyrGluProSerTrpGluGluGlnProMetProGluThrProLysTrpPh 125
|||.....|
466 .....AGCCAGGTCTCTGAGTTCGCCGAGTGCACAC 494
125 eHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSerSerSerLysP 142
|||.....|
495 CCAGGTC.....TACACAAAGGACATAGGCA 530
142 heValThrArgValThrThrProAlaGlnArgGlyLysValPro 156
|||.....|
521 GAAATGTGACCATGGAATGCCCTTCAAAAGGGAGGAATGTCC 564

```

seq\_name: N\_Geneseq\_36:V15422

seq\_documentation\_block:

ID V15422 standard; DNA; 29392 BP.

AC V15422;

DT 11-JUN-1998 (first entry)

DE Mouse poly Ig receptor protein gene.

KW Mouse; poly Ig receptor protein; pIgR protein; pIg; deficiency;

KW knockout mouse; disease model; ds.

OS Mus sp.

PN J10057066-A.

PD 03-MAR-1998.

PF 19-AUG-1996; 217154.

PR 19-AUG-1996; JP-217154.

PA (HONS ) YAKULT HONSHA KK.

DR WPI; 98-254323/23.

PT Mouse pIg receptor protein gene - used for preparing gene knockout

PT mice, useful for study of human poly Ig receptor protein deficiency

PS Claim 1; Page 4-14; 18pp; Japanese.

CC The present sequence represents the mouse poly Ig receptor protein

CC gene, which has a 29392 bp sequence. The new gene can be used to

CC produce a gene knockout mouse, useful as a disease model of human

CC poly Ig receptor protein deficiency.

SQ Sequence 29392 BP; 8318 A; 6747 C; 6514 G; 7813 T;

seq_name	N_Geneseq_36:T31291	seq_documentation_block:
US-09-135-238B-2_COPY_18_253 x V15422	..	Alignment seg 1/1 to: V15422 from: 1 to: 29392
alignment_scores:		
Quality:	185.50	Length: 164
Ratio:	1.855	Gaps: 6
Percent Similarity:	60.976	Percent Identity: 30.488
alignment_block:		
US-09-135-238B-2_COPY_18_253 x T31291	..	
Align seg 1/1 to: T31291	from: 1 to: 3269	
13 GlyGlySerValThrLeuLysCysProLeuProGlu.....Me 25		
170 GGTAACTCGGTCTCCATCACTGCTACTACCCAGACACCTCTGTCAACCG 219		
25 tHsValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 42		
220 GCACACCCGGAATACTGGTGGCCGACAGAGGCC...AACGGCTACTGCG 266		
42 lyThrValValSerThrThrAsnPhelIleLysAlaGluTyrLysGlyArg 58		
267 CAACCTCATCTCTTCAATGGCTACCTCTCGAAGGAGTATTTCAGGCAGA 316		
59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75		
317 GCCAGCCTCATCACTTCCAGAGATAGACACATTTGTGATTAACTGTC 366		
75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 92		
367 ACATCTCACCCAGGAGACACTGGGAGCTCAAGTGTGCTGGGT...A 413		
92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 108		
414 CCATCAACCGAGGCTGTTTCGATGTCAGCCTGGAGGTC.....454		
109 TyrGluProSerThrGluGlnGluInProMetProGluThrProLysTrpPh 125		
455 .....AGCCAGGTCTCTGAGTCCCAATGACAC 483		
125 eHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSerSerSerLysP 142		
484 CCATGTC.....TACACAAAGG 500		
142 heValThrArgValThrProAlaGlnArgGlyLysValProVal 158		
501 ACATAGGAGCAACTGTCACCATCGAATGCCGCTTTCAAGAGGGGAATGCT 550		
159 HisHisSer.SerProThrThrGlnIleThrHisArgPro 171		
551 CATAGCAAGAAATCCCTGTGTAAGAAGAGAGAGAGGCGCT 590		
seq_name: N_Geneseq_36:X07407		
seq_documentation_block:		
ID X07407 standard; cDNA to mRNA; 1839 BP.		
AC X07407;		
DT 08-JUN-1999 (first entry)		
DE Human secretory Immunoglobulin A component gene.		
KW Immunoglobulin A; secretory; component; IgA; human; treatment;		
KW prevention; infection; HIV; AIDS; cold; flu; virus; gene;		
KW human immunodeficiency virus; respiratory syncytial virus; ss.		
OS Homo sapiens.		
FT Key	Location/Qualifiers	
FT CDS	1..1839	
FT	/tag= a	
FT	/product= secretory component	

OM of: US-09-135-238B-2 to: N\_Geneseq\_36:\* out\_format : pfs  
 Date: Sep 12, 2000 6:55 AM  
 About: Results were produced by the GenCore software, version 4.5,  
 Copyright (c) 1993-2000 CompuGen Ltd.

## Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
 -O=/cgn2.1/USPTO\_spool/US09135238/runat\_24082000\_091324.17918/app\_query.fasta\_1.1282  
 -DB=N\_Geneseq\_36 -QFMT=fastap -SURFIX=p2n.rng -GAPOP=12.000  
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
 -GAPOP=4.500 -GAPEXT=0.050 -GAPOP=10.000 -GAPEXT=0.500  
 -GAPOP=6.000 -GAPEXT=7.000 -GAPOP=10.000 -GAPEXT=0.500  
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blsum62  
 -TRANS=human4.0.cdi -LIST=45 -DCCALIGN=200 -THR\_SCORE=pct  
 -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
 -NORM-ext -MINLEN=0 -MAXLEN=100000  
 -USER=US09135238 -CGN1\_1.123 -NCPU=6 -ICPU=3 -LONGLOG -NO\_XLPPX  
 -WAIT -THREADS=1

## Search information block:

Query: US-09-135-238B-2  
 Query length: 390  
 Database: N\_Geneseq\_36:\*  
 Database sequences: 311585  
 Database length: 125096042  
 Search time (sec): 121.040000

## score\_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
N_Geneseq_36:X28178	+ 2047.00	1229.34	5.3e-111	2040	Human PIGRL-1 coding sequence.
N_Geneseq_36:X28179	+ 1125.00	1174.09	8.5e-58	1047	Human PIGRL-1 coding sequence.
N_Geneseq_36:T31220	+ 198.00	199.21	0.0017	3095	Mouse poly-immunoglobulin rece
N_Geneseq_36:T31221	+ 188.00	170.27	0.0697	29392	Mouse poly Ig receptor protei
N_Geneseq_36:T31222	+ 187.50	192.24	0.0042	1911	CDNA for human immunity relat
N_Geneseq_36:T31223	+ 185.50	185.74	0.0096	3269	Rat poly-immunoglobulin recept
N_Geneseq_36:T31224	+ 185.00	186.53	0.0059	1936	Secretory component DNA fragme
N_Geneseq_36:T31225	+ 182.00	189.82	0.0083	1839	Human secretory immunoglobulin
N_Geneseq_36:T31226	+ 182.00	186.01	0.0093	2031	Secretory component DNA fragme
N_Geneseq_36:T31227	+ 182.00	183.02	0.0136	2919	Human poly-immunoglobulin rece
N_Geneseq_36:T31228	+ 182.00	181.23	0.0171	3630	Bovine poly-immunoglobulin rece
N_Geneseq_36:T31229	+ 171.00	174.33	0.0414	2084	Rat kidney injury related mole
N_Geneseq_36:T31230	+ 171.00	172.62	0.0515	2566	Rat kidney injury related mole
N_Geneseq_36:T31231	+ 166.50	170.51	0.0676	1876	Clone 1 for truncated poly Ig
N_Geneseq_36:T31232	+ 166.50	165.34	0.1312	3517	Rabbit poly-immunoglobulin rec
N_Geneseq_36:T31233	+ 129.50	167.34	0.1015	1137	Human PIGR-2 coding sequence.
N_Geneseq_36:T31234	+ 143.50	148.11	1.20	1876	Clone 2 for truncated poly Ig
N_Geneseq_36:T31235	+ 143.50	133.93	7.37	8690	Plasmid pD17-hum14.H1. Inhibit
N_Geneseq_36:T31236	+ 142.00	150.90	0.8361	914	Human EST R2810 DNA. Complexes
N_Geneseq_36:T31237	+ 139.50	130.12	12.02	8321	Plasmid expressing hBR96-A. In
N_Geneseq_36:T31238	+ 138.50	146.63	1.44	985	Mycobacterium tuberculosis anti
N_Geneseq_36:T31239	+ 138.50	146.63	1.44	985	M. tuberculosis immunogenic pol
N_Geneseq_36:T31240	+ 136.50	149.12	1.05	565	EST clone EN260. New polynucle
N_Geneseq_36:T31241	+ 136.50	140.14	3.32	1683	Human dUTPase gene 5' region.
N_Geneseq_36:T31242	+ 136.00	123.79	27.06	11529	Plasmid pAH4602. Antibody con
N_Geneseq_36:T31243	+ 135.00	123.35	28.60	10704	Plasmid pAH4625. Antibody con
N_Geneseq_36:T31244	+ 131.00	118.16	55.71	12132	Plasmid pAH4807. Antibody con
N_Geneseq_36:T31245	+ 130.00	133.70	7.59	1617	Antibody D heavy chain. Prodn.
N_Geneseq_36:T31246	+ 129.50	134.18	7.13	1431	Primate anti-human B7.1 ant
N_Geneseq_36:T31247	+ 129.50	134.18	7.13	1431	Macaque primatized 7C10 heavy
N_Geneseq_36:T31248	+ 127.00	128.05	15.66	2197	IS1096 transposon. Leucine aux
N_Geneseq_36:T31249	+ 126.50	120.88	39.26	4266	Plasmid pHC605 nucleotide sequ
N_Geneseq_36:T31250	+ 126.50	114.03	94.63	11336	Approximate nucleotide sequen
N_Geneseq_36:T31251	+ 125.50	129.22	13.48	1576	Human anti-HBs heavy chain. Hu
N_Geneseq_36:T31252	+ 125.50	122.89	30.35	3400	3F4 human G2/G4 chimeric antib
N_Geneseq_36:T31253	+ 125.50	116.64	67.64	7266	Human calcium channel a1B subu
N_Geneseq_36:T31254	+ 124.50	116.45	69.37	6557	Anti-IgE VH expression vector.
N_Geneseq_36:T31255	+ 124.50	108.70	187.39	16812	HSV-2 strain S85 Contig ID 12
N_Geneseq_36:T31256	+ 124.00	125.90	20.63	1351	C. acidovorans gamma-lactamase
N_Geneseq_36:T31257	+ 124.00	124.32	25.27	2364	CD4-specific CDR-grafted heavy

N\_Geneseq\_36:T50962 + 124.00 115.30 80.34 7073 ! TF8-569 CDR-grafted heavy c  
 N\_Geneseq\_36:T57472 + 123.50 125.47 21.80 1929 ! Sorghum bicolor (L.) Moench  
 N\_Geneseq\_36:T96035 + 123.00 125.54 21.60 1795 ! Human kidney injury related  
 N\_Geneseq\_36:T72685 + 123.00 113.19 105.27 8051 ! Sugar biosynthesis gene clu  
 N\_Geneseq\_36:X25774 + 123.00 113.08 106.78 8160 ! S.erythraea erythromycin-sy

seq\_name : N\_Geneseq\_36:X28178

## seq\_documentation\_block:

ID X28178 standard; cDNA; 2040 BP.  
 AC X28178;  
 DT 16-JUN-1999 (first entry)  
 DE Human PIGRL-1 coding sequence.  
 KW PIGRL-1; human; autoimmune disease; hyper-IgM Immunodeficiency; HIM;  
 KW X-linked Severe Combined Immunodeficiency; XSCID; IgA deficiency;  
 KW diagnosis; therapy; ss.  
 OS Homo sapiens.  
 PN EP-905238-A2.  
 PD 31-MAR-1999.  
 PF 14-AUG-1998; 306487.  
 PR 30-OCT-1997; US-961564.  
 PR 25-AUG-1997; US-056935.  
 PA (SMIK) SMITHKLINE BEECHAM CORP.  
 PI Sweet RW, Truneh A, Wu S;  
 DR WPI: 99-192666/17.  
 DR P-PSDB: Y05001.  
 PT New polypeptides encoding PIGRL-1 useful for treating diseases such  
 as X-linked Severe Combined Immunodeficiency  
 PS Claim 4; Page 18-19; 28pp; English.  
 CC This sequence encodes the human PIGRL-1 protein of the invention.  
 CC Autoimmune diseases involving altered expression or activity of PIGRL-1  
 CC may include Hyper-IgM Immunodeficiency (HIM), X-linked Severe Combined  
 CC Immunodeficiency (XSCID) and IgA deficiency. These diseases can be  
 CC diagnosed or susceptibility to them predicted by: (1) determining whether  
 CC there is a mutation in the genomic copy of the gene encoding PIGRL-1; or  
 CC (2) measuring the amount of PIGRL-1 in a sample derived from the patient.  
 CC Patients deficient in PIGRL-1 can be treated by administering either the  
 CC PIGRL-1 DNA or its complement or an agonist of PIGRL-1 to the patient.  
 CC Patients with excessive expression or activity of PIGRL-1 can be treated  
 CC by administering an antagonist of PIGRL-1, an antisense nucleic acid  
 CC molecule which inhibits the expression of PIGRL-1 or administering  
 CC sufficient PIGRL-1 to compete with the endogenous activity. PIGRL-1 can  
 CC be used to identify its agonists by contacting a cell expressing PIGRL-1  
 CC with a candidate compound in the presence of a signal system and noting  
 CC the candidate as an agonist if a signal is produced. The same method can  
 CC be used to identify antagonists of PIGRL-1 but the presence of an  
 CC antagonist is indicated by a decrease in production of the signal.  
 CC Antibodies against PIGRL-1 may be used to isolate or identify clones  
 CC expressing PIGRL-1. Polynucleotides encoding PIGRL-1 may be used to  
 CC identify chromosomal mutations in the gene encoding PIGRL-1 in patients.  
 CC This information may then be correlated with the incidence of autoimmune  
 CC disease in those patients to identify whether the mutation causes the  
 CC disease.

## alignment\_scores:

Quality: 2047.00 Length: 390  
 Ratio: 5.262 Gaps: 0  
 Percent Similarity: 99.744 Percent Identity: 99.744

## alignment\_block:

US-09-135-238B-2 x X28178  
 Align seg 1/1 to: X28178 from: 1 to: 2040  
 1 MetaspargtrpLeuprpProLeuYrPheLeuProValSerGlyAlaLe 17  
 146 ATGGACCTCTGGCTTTGGCCACTTCTCTCTCCATGATCAGGGCCCT 195  
 17 ATGATLeuProGluValValGluGluGluGlyGlySerValT 34  
 196 GAGGATCTCTCCCAAGAGTAAGAGTAGAGGGGAGCTGGCGGATCAGTTA 245

34 hrileLysCysProLeuProGluMetHisValArgIleLeuLeuCysArg 50  
 246 CCATCAAGTGGCCACTTCTCTGAAATGCATGTGAGGATATATCTGTGCGG 295

51 GluMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsp 67  
 296 GAGATGGCTGGATCTGGACATGTGTACCTGGTATCCACACCACTT 345

67 eileLysAlaGluThrLysGlyArgValThrLeuLysGlnThrProArgL 84  
 346 CATCAAGGAGCAATACAAAGGGCGGAGTTACTCTGAAGCAATACCCACGCA 395

84 ysAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerGly 100  
 396 AGAATCTGTCTAGTGGAGTAACACAGCTGACAGAAAGTGACAGCGGA 445

101 ValTyAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGln 117  
 446 GTCTATGCTTGGGAGCGGCGATGAACACAGACCGGGGAAAGACCCAGAA 495

117 sValThrLeuAsnValHisSerGluThrGluProSerThrGluGlnP 134  
 496 AGTCACCCCTGAATCTCCACAGTGAATACGAGCCATCATGGGAAGACGAGC 545

134 roMetProGluThrProLysThrPheHisLeuProTyLeuPheGlnMet 150  
 546 CAATGCTGAGACTCCAAATGGTTTCATCTGCGCTATTTGTTCCAGATG 595

151 ProLysThrAlaSerSerSerLysPheValThrArgValThrProAl 167  
 596 CTCTCATGTCAGTCTCTCAAAATTCGTAACACAGAGTTTACACACACGAGC 645

167 aGlnArgGlyLysValProProValHisHisSerSerProThrThrGln 184  
 646 TCAAGGGGCAAGGTCCTCCAGTTCACCACTCTCTCCCGCCACCCACAA 695

184 leThrHisArgProArgValSerArgAlaSerSerValAlaGlyAsp 200  
 696 TCACCCACCGCCTCGAGTGCAGAGCATCTTCAGTAGCAGGTGACAAAG 745

201 ProArgThrPheLeuProSerThrThrAlaSerLysIleSerAlaLeuG 217  
 746 CCCGGAACCTTCTGCCATCCACTACAGCCTCAAAATCTCAGCTCTGGA 795

217 uGlyLeuLeuLysProGlnThrProSerTyAsnHisThrArgLeuH 234  
 796 GGGGCTCTCAGCCCGCCAGAGCCGAGCTCAACACACACACAGGCTGC 845

234 isArgGlnArgAlaLeuAspTyArgLysGlnSerGlyArgGluGln 250  
 846 ACAGGCAGAGACACTGAGTATGGCTCAGCTCTGGGAGGGAAGGCCAA 895

251 GlyPheHisLeuLeuLeuProThrIleLeuGlyLeuPheLeuAlaLe 267  
 896 GGATTCACATCTCATCCCGACCATCTGGGCGCTTTCTGCTGCACT 945

267 uLeuGlyLeuValValLysArgAlaValGluArgArgGlyAlaLeuSer 284  
 946 TCTGGGCTGTGGTGAAGAGGCGCTTGAAGAGGAGGAAAGCCCTCTCCA 995

284 rArgAlaArgArgLeuAlaValArgMetArgAlaLeuGluSerSerGln 300  
 996 GSGCGGCGCGCGACTGGCGGTGAGGATGCGCGCTTGAGAGCTCCCG 1045

301 ArgProArgGlySerProArgProArgSerGlnAsnAsnIleTySerAl 317  
 1046 AGGCCCCGGGGTCCGCGCGAGCCGCTCCCAAAACACATCTACAGCGC 1095

317 aCysProArgArgAlaArgGlyAlaAspAlaAlaGlyThrGlyGluAlaP 334  
 1096 CTGCGCGCGCGCGCTCTGTGAGCGGACGCTGACAGGACAGGGAGGCC 1145

334 roValProGlyProGlyAlaProLeuProProAlaProLeuGlnValSer 350  
 1146 CCGTCCCGCGCGGAGCGCGCTTGCCTCCCGCGCGCTGCGAGGTGCT 1195

351 GluSerProTyrLeuHisAlaProSerLeuLysThrSerCysGluTyVa 367  
 1196 GAATCTCCCTGGCTCCATGCCCATCTCTGAGACAGCTGTGAATAGT 1245

367 lserLeuTyHisGlnProAlaAlaMetMetGluAspSerAspSerAsp 384  
 1246 GAGCTCTACCCAGCTGCCCATCTGATGAGGAGCAGTGATTCAGATG 1295

384 spTyIleAsnValProAla 390  
 1296 ACTACATCAATGTTCTCTGCC 1315

seq\_name: N\_Geneseq\_36:X28179

seq\_documentation\_block:  
 ID X28179 standard; CDNA; 1047 BP.  
 AC X28179;  
 DT 16-JUN-1999 (first entry)  
 DE Human PIGRL-1 coding sequence fragment.  
 KW PIGRL-1; human; autoimmune disease; hyper-IgM Immunodeficiency; HIM;  
 KW X-linked Severe Combined Immunodeficiency; XSCID; IgA deficiency;  
 KW diagnosis; therapy; ss.  
 OS Homo sapiens.  
 PN EP-905238-A2.  
 PD 31-MAR-1999.  
 PF 14-AUG-1998; 306487.  
 PR 30-OCT-1997; US-961564.  
 PR 25-AUG-1997; US-056935.  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 PI Sweet RW, Truneh A, Wu S;  
 DR WPI: 99-192666/17.  
 DR P-PSDB: Y05002.  
 PT New polypeptides encoding PIGRL-1 useful for treating diseases such  
 as X-linked Severe Combined Immunodeficiency  
 PT Disclosure: Page 8; 26pp; English.  
 CC This sequence encodes the human PIGRL-1 protein of the invention.  
 CC Autoimmune diseases involving altered expression or activity of PIGRL-1  
 CC may include Hyper-IgM Immunodeficiency (HIM), X-linked Severe Combined  
 CC Immunodeficiency (XSCID) and IgA deficiency. These diseases can be  
 CC diagnosed or susceptibility to them predicted by: (1) determining whether  
 CC there is a mutation in the genomic copy of the gene encoding PIGRL-1; or  
 CC (2) measuring the amount of PIGRL-1 in a sample derived from the patient.  
 CC Patients deficient in PIGRL-1 can be treated by administering either the  
 CC PIGRL-1 DNA or its complement or an agonist of PIGRL-1 to the patient.  
 CC Patients with excessive expression or activity of PIGRL-1 can be treated  
 CC by administering an antagonist of PIGRL-1, an antisense nucleic acid  
 CC molecule which inhibits the expression of PIGRL-1 or administering  
 CC sufficient PIGRL-1 to compete with the endogenous activity. PIGRL-1 can  
 CC be used to identify its agonists by contacting a cell expressing PIGRL-1  
 CC with a candidate compound in the presence of a signal system and noting  
 CC the candidate as an agonist if a signal is produced. The same method can  
 CC be used to identify antagonists of PIGRL-1 but the presence of an  
 CC antagonist is indicated by a decrease in production of the signal.  
 CC Antibodies against PIGRL-1 may be used to isolate or identify clones  
 CC expressing PIGRL-1. Polynucleotides encoding PIGRL-1 may be used to  
 CC identify chromosomal mutations in the gene encoding PIGRL-1 in patients.  
 CC This information may then be correlated with the incidence of autoimmune  
 CC disease in those patients to identify whether the mutation causes the  
 CC disease.  
 SQ Sequence 1047 BP; 255 A; 281 C; 276 G; 223 T;

alignment\_scores:  
 Quality: 1125.00 Length: 275  
 Ratio: 4.429 Gaps: 12  
 Percent Similarity: 92.364 Percent Identity: 90.182

alignment\_block:  
 US-09-135-238B-2 x X28179

Align seg 1/1 to: X28179 from: 1 to: 1047

```
1 MetAspArgTrpLeuProLeuValPheLeuProValSerGlyAlaLeu 17
|||||
150 ATGAGCTCTGGCTTTGGCCACATTTACTTCCTGCCAGTATCAGGGCCCT 199
17 uArgileLeuProGluValLysValGluGlyLeuGlyGlySerValT 34
|||||
200 GAGGATCTCCCAAGAGTAAGGTAGAGGGGAGCTGGCGGATCAGTTA 249
34 hrileLysCysProLeuProGluMetHisValArgileTyrLeuCysArg 50
|||||
250 CCATCAAGTGCCACATTCCTGAAATCATGTGAGGATATATCTGCGCG 299
51 GluMetAlaGlySerGlyThrCysGlyThrValValSerThrAsnPh 67
|||||
300 GAGATGCTGGATCTGGACATGTGTACTCGGTATCCACCACCAACTT 349
67 eileLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProArgL 84
|||||
350 CATCAAGGCAGATAACAAGGCGGAGTTACTCTGAAGCAATACCCACGCA 399
84 ysAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerGly 100
|||||
400 AGAATCTGTCTCTAGTGGAGGTAAACACAGCTGACAGAAAGTGACAGCGGA 449
101 ValTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLy 117
|||||
450 GTCATGCTCGGGAGC.GGCATGAACACAGACCGGGGAAGACCCAGAA 498
117 sValThrLeuAsnValHisSerSerGlyTyrGluProSerTrpGluGlnP 134
|||||
499 AGTCACCTGTAATGTCCACAGTGAATACGAGCCATCATGGGAAGAGCAGC 548
134 rOmetProGluThrProLysTrpPheHisLeuProTyrLeuPheGlnMet 150
|||||
549 CAATGCTGAGACTCAAATGGTTTCATCTGCTCCCTATTGTTCAGATG 598
151 ProAlaTyrAlaSerSerSerLysPheValThr..ArgValThrThrPro 166
|||||
599 CCTGATATCGCGGTCTTCCACATTCGTAAACCGCAGAGTTACCAACCA 648
167 Ala.GlnArgGlyLysValProValHisHisSerSerProThrThrG 183
|||||
649 GCTTCAAAAGGGCAAGGTCCCTCCAGTTCACCACTCTCCGCCACCCACC 698
183 Inile..ThrHisArgPro.ArgValSerArgAlaSerSerValAlaGlyA 199
|||||
699 AAATTCAACCCACCGCCCTTCGAGTGTNCAGAGCATCTTCAGTAGCAGGTG 748
199 sPLysProArg..ThrPheLeuProSerThrThrAlaSerLysIleSerAl 215
|||||
749 ACAAGCCCGCAAACTTCTCGCATCCACTACAGCCTCAAAATCTCAGC 798
215 aLeuGluGlyLeuLeuLys....ProGluThrProSerTyrAsnHisHist 231
|||||
799 YCTGAAGGGTGCTTCAAGCCCGCAAGAGCCCGCAGCTACAAACAN.CACA 847
231 hrArgLeuHisArgGlnArgAlaLeuAsp...Tyr.GlySerGlnSerG 246
|||||
848 CCAGGCTCCACAGGCAGAGACACTGGATATCTATGGNTCCACAGTCTGG 897
246 YArgGluGly.....GlnGlyPhe...HisIleLeuIlePro.ThrIle 259
|||||
898 G...GAGGGGAANGNCCAAGGATTTNACATTCCTGATTCCTCCGCGACCATC 944
260 LeuGlyLeuPhe 263
|||||
945 NPTGGGGCCTTT 956
```

seq\_name: N\_Geneseq\_36:T31290

seq\_documentation\_block:

```
ID T31290 standard; CDNA; 3095 BP.
AC T31290;
DT 24-FEB-1997 (first entry)
DE Mouse poly-immunoglobulin receptor, cDNA.
KW Mouse; immunoglobulin; receptor; protection protein; mutants;
KW heavy chain; antigen binding domain; protection; pathogen;
KW mucosal; environment; gastrointestinal; passive; immunisation;
KW Guy's 13 antibody; prevention; dental caries; Streptococcus;
KW poly; sorbinus; murine; ss.
OS Mus musculus.
FH Key Location/Qualifiers
FT 85..2400
FT cds /*tag= a
FT
PD WO9621012-Al.
PN 11-JUL-1996.
PF 27-DEC-1995; U16889.
PR 30-DEC-1994; US-367395.
PR 04-MAY-1995; US-434000.
PA (PLAN-) PLANT BIOTECHNOLOGY INC.
PA (UNKE-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.
PA (PLAN-) PLANET BIOTECHNOLOGY INC.
PI Hiatt AC, Lehner T, Ma JKC;
DR WPI; 96-333987/33.
DR P-PSDB; W03180.
PT Immunoglobulin and protection protein complex and its prodn. in
PT plants - useful for passive immunisation against mucosal antigens,
PT esp. against S. mutans and S. sorbinus to prevent dental caries
PS Disclosure: Pages 117-121; 152pp; English.
CC The present sequence encodes the mouse poly-immunoglobulin (Ig)
CC receptor, a portion of which corresp. to residues 1-627, pref.
CC 1-606, or esp. residues 13-45, 1-120, 110-230, 210-340, 320-450,
CC 440-550, 550-606 or 550-627 comprises a protection protein (PP).
CC The Ig of the invention comprises a pp as above in association with
CC an Ig derived heavy chain, having at least a portion of an antigen
CC (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.
CC gastrointestinal, environments, therefore enhancing its
CC effectiveness in passively immunising animals against mucosal
CC pathogens. The Ag binding domain is specifically derived from the
CC Guy's 13 antibody, and the Ig can be used to prevent dental caries
CC by binding, e.g. Streptococcus mutans serotypes c, e and f, or
CC S. sorbinus serotypes d and g.
SQ Sequence 3095 BP; 861 A; 796 C; 784 G; 654 T;
```

alignment\_scores:

Quality:	198.00	Length:	168
Ratio:	1.980	Gaps:	5
Percent Similarity:	59.524	Percent Identity:	31.548

alignment\_block:

US-09-135-238B-2 x T31290 ..

Align seg 1/1 to: T31290 from: 1 to: 3095

```
10 PheLeuProValSerGlyAlaLeuArgileLeuProGluValLysValG 26
|||
121 TTTTCAGGGTCTCCACAAAAGCCCATTTGGTCCCCAGAGGTGAG 170
26 uGlyGluLeuGlyGlySerValThrIleLysCysProLeuProGlu... 41
|||
171 TAGTATAGAGGCGACTCTGTTTCCATCACGTCTACTACCCAGACACT 220
42 .....MethisValArgileTyrLeuCysArgGluMetAlaGlySer 55
|||
221 CTGTCAAACCGGCACACCCGGAATACTGTCGCCGACAAGGAGCC...AGC 267
56 GlyThrCysGlyThrValValSerThrThrAsnPheIleLysAlaGluT 72
|||
268 GGCATGTGCACAACGCTCATCTCTTCAATGGCTACCTCTCCAAGAGTA 317
72 rLysGlyArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeu 89
|||
```

```

318 TTCAGGCGAGCCAACTCATCACTCCAGAGAACACACATTTGTGA 367
89  aGlulValThrGlnLeuThrGluSerAspSerGlyValTyAlaCysGly 105
      ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
368  TTAACATTGACGAGCTCAACAGGACGACACTGGGAGCTACAAGTGGC 417
      |||  ::::: ||||| ::::: ||||| ::::: |||||
106  AlagLyMetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnVa 122 a
      |||  ::::: ||||| ::::: ||||| ::::: |||||
418  CTGGGT...ACCAAGTAACGAGGCTGTCTTCGATGTCAGCTGGAGGT 464
122  lHisSerGluTyGluProSerTrpGluGluGlnProMetProGluThrP 139
      |  ::::: ||||| ::::: ||||| ::::: |||||
465  C.....AGCCAGGTCTCTGAGTTGC 484
139  rOLysTrpPheHisLeuProTyLeuPheGlnMetProAlaTyAlaSer 155
      ||::: ||||: ||||: ||||: ||||:
485  CGAGTGACACCCAGCTC.....TACACAAAG 510
156  SerSerLysPheValThrArgValThrThrProAlaGlnArgGlyLysVa 172
      :::  :::  |||  ::::: ||||| ::::: |||||
511  GACATAGCGAGAAATGTGACCATTAATGCCCTTTCAAAGGAGGAGATGT 560
172  lPro 173
561  TCCC 564
seq_name: N_Geneseq_36:V15422
seq_documentation_block:
ID V15422 standard; DNA; 29392 BP.
AC V15422;
DT 11-JUN-1998 (first entry)
DE Mouse poly Ig receptor protein gene.
KW Mouse; poly Ig receptor protein; pIgr protein; pIg; deficiency;
KW knockout mouse; disease model; ds.
OS Mus sp.
PN J10057066-A.
PD 03-MAR-1998.
PF 19-AUG-1996; 217154.
PR 19-AUG-1996; JP-217154.
PA (HONS ) YAKULT HONSHA KK.
DR WPI; 98-254323/23.
PT Mouse pig receptor protein gene - used for preparing gene knockout
PT mice, useful for study of human poly Ig receptor protein deficiency
PS Claim 1; Page 4-14; 18pp; Japanese.
CC The present sequence represents the mouse poly Ig receptor protein
CC gene, which has a 29392 bp sequence. The new gene can be used to
CC produce a gene knockout mouse, useful as a disease model of human
CC poly Ig receptor protein deficiency.
SQ Sequence 29392 BP; 8318 A; 6747 C; 6514 G; 7813 T;

alignment_scores:
Quality: 188.00 Length: 97
Ratio: 2.648 Gaps: 3
Percent Similarity: 73.196 Percent Identity: 40.206

alignment_block:
US-09-135-238b-2 x V15422 ..
Align seg 1/1 to: V15422 from: 1 to: 29392
30 GlySerValThrIleLysCysProLeuProGlu.....Me 42
||| ||||| ::::: |||||
18245 GCGGACTGTGTTCCATCAGTCTACTACCCAGACACACTCTGTCAACG 18294
42 tHisValArgIleTyLeuCysArgGluMetAlaGlySerGlyThrCysG 59
||| ||||| ::::: ||||| ::::: |||||
18295 GCACACCCGGAATACTGGTGGCAGACAGGAGCC....AGCGGATGTGCA 18341
59 lyThrValValSerThrThrAsnPheIleLysAlaGluTyLysGlyArg 75
||| ::::: ||||| ::::: ||||| ::::: |||||
18342 CAAGGCTCATCTCTCAAAATGGCTACCTCTCAAGGAGGATTCAGGCAGA 18391

```

```

76 ValThrLeuLysGlnTyProArgLysAsnLeuPheLeuValGluValTh 92
      ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
18392 GCCAACTCATCACTCCAGAGAACACACATTTGTGATTAACTGA 18441
92 rGlnLeuThrGluSerAspSerGlyValTyAlaCysGlyAlaGlyMeta 109
      ||||| ::::: ||||| ::::: ||||| ::::: |||||
18442 GCAGCTCAACCCAGGACGACACTGGGAGCTACAAGTGGCGTGGGT...A 18488
109 snThrAspArgGlyLysThrGlnLysValThrLeuAsnVal 122
      ::::: ||||| ::::: ||||| ::::: |||||
18489 CCAGTAACGAGGCTGTCTTCGATGTCAGCTGGAGGTC 18529
seq_name: N_Geneseq_36:V20383
seq_documentation_block:
ID V20383 standard; cDNA to mRNA; 1911 BP.
AC V20383;
DT 26-JUN-1998 (first entry)
DE cDNA for human immunity related factor.
KW lymph node; human; immunity related factor; research; treatment;
KW immune disease; infectious disease; ds.
OS Homo sapiens.
FH Key
FT CDS
      Location/Qualifiers
      98..1696
      /*tag= a
      98..145
      /*tag= b
      146..1693
      /*tag= c
      /product= immunity_related_factor
      J10072495-A.
      17-MAR-1998.
      PD 11-JUN-1997; 153218.
      PF 13-JUN-1996; JP-152362.
      PA (ASAH ) ASahi KASEI KOGYO KK.
      DR WPI; 98-234766/21.
      PR P-FSDB; W50033.
      PT Immunity related factor - useful in the treatment of immune related
      PT and infectious diseases
      PS Claim 15; Pages 18-20; 21pp; Japanese.
      CC The present sequence encodes a lymph node derived human immunity
      CC related factor, which can be used to research and treat immune and
      CC infectious diseases.
      SQ Sequence 1911 BP; 490 A; 541 C; 525 G; 355 T;

alignment_scores:
Quality: 187.50 Length: 575
Ratio: 0.906 Gaps: 20
Percent Similarity: 36.000 Percent Identity: 18.957

alignment_block:
US-09-135-238b-2 x V20383 ..
Align seg 1/1 to: V20383 from: 1 to: 1911
3 ArgTrpLeuTrpProLeuTyPheLeuProValSerGlyAlaLeuArgI1 19
||| ||||| ::::: |||||
170 AGATGGCTGTGGAG...GGCTCTCTCCCTCCAGGACCCATCTCCGGGC 216
19 eLeuProGluValLys..... 24
::: :::::
217 CATGGGAACACTCAGGCTCTCTCGCCCTCTCTCGCGGCGGAGAGAGCT 266
25 .....ValGluGlyGlu 28
||| ::::: |||||
267 CTTTTCAGCTCCAAATTCATTGAAGGGCTCAAGGCTGTGTCTCAGGGAG 316
29 LeuGlyGlySerValThrIleLysCysProLeuPro.....G1 41
||| ::::: ||||| ::::: |||||
317 CTGGAGGAGCTGTCTCACCATTCAGTGCCATTATGCCCTCATCTGTCAA 366

```

```

41 uMethHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrC 58
   :   |||   |||   |||   |||   |||   |||   |||   |||
367 CAGGCACCAAGAGAAAGTACTGTCGCGCTGGGGCCCCCAAGATGGATCT 416
   ||   |||   |||   |||   |||   |||   |||   |||   |||
58 yGlyThrValValSerThrThrAsnPheIleLysAlaGluTyrLysGly 74
   ||   |||   |||   |||   |||   |||   |||   |||   |||
417 GCACAGCAATGTCGTCACCAACCAAGTATATCTCACCATCGCTATCGTAC 466
   ||   |||   |||   |||   |||   |||   |||   |||   |||
75 ArgValThrLeuLysGlyTyrProArgLysAsnLeuPheLeuGluVa 91
   ||   |||   |||   |||   |||   |||   |||   |||   |||
467 CGTGTGGCCCTCAGAGATTTCCACAGAGAGCTTGTGTGGTGAGGCT 516
   ||   |||   |||   |||   |||   |||   |||   |||   |||
91 lThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyM 108
   ||   |||   |||   |||   |||   |||   |||   |||   |||
517 GTCCCAACTGTCCCGGATGATCGGATGCTACCTCTGCGGCATTGAA 566
   ||   |||   |||   |||   |||   |||   |||   |||   |||
108 eTAsnThrAspArg..... 112
   ||   |||   |||   |||   |||   |||   |||   |||   |||
567 GTGAACAACAATGCTGTCTTAAAGCATGAATCTGACCATCTCGCAGGT 616
   ||   |||   |||   |||   |||   |||   |||   |||   |||
112 ..... 112
617 CCGGCCAGCACCCCTCCACAGCCACTCCAGCTGCTGGGGAGCTCACCAT 666
   ||   |||   |||   |||   |||   |||   |||   |||   |||
113 .....GlyL 114
   ||   |||   |||   |||   |||   |||   |||   |||   |||
667 GAGATCCTTGAACAGCGCTCTCCAGTGGCCACAGATGGACCCAGAA 716
   ||   |||   |||   |||   |||   |||   |||   |||   |||
114 ysThrGln..... 116
   ||   |||   |||   |||   |||   |||   |||   |||   |||
717 CCACCCAGACTTAGCAGAGGGACAGCATGGGACACAGTTGCTTCCACT 766
   ||   |||   |||   |||   |||   |||   |||   |||   |||
117 .....LysValThrLeuAsnValHisSerGluTyrGluPro.. 128
   ||   |||   |||   |||   |||   |||   |||   |||   |||
767 CCAGGAACCAAGACAGACTACAGCTTCAGCTCAGGGAAGACCAACCCAGG 816
   ||   |||   |||   |||   |||   |||   |||   |||   |||
129 .....SerTrpGluGlu..... 132
   ||   |||   |||   |||   |||   |||   |||   |||   |||
817 AGCAACCAAGCCAGCAGCTCCAGGGACAGGCAGCTGGCGCAGGGTCTG 866
   ||   |||   |||   |||   |||   |||   |||   |||   |||
133 .....GlnProMetProGluThrProLysTyrPheHisLeuPro 145
   ||   |||   |||   |||   |||   |||   |||   |||   |||
867 TCAAGCACTGCTCCGATCCAGAGTCCA..... 898
   ||   |||   |||   |||   |||   |||   |||   |||   |||
146 TyrLeuPheGlnMetProAlaTyrAlaSerSerLysPheValThrAr 162
   ||   |||   |||   |||   |||   |||   |||   |||   |||
899 .....CCITCAAAGACAGACAGCATGTCCAA 924
   ||   |||   |||   |||   |||   |||   |||   |||   |||
162 gValThrThrProAlaGlnArgGlyLysValProProValHisHisSerS 179
   ||   |||   |||   |||   |||   |||   |||   |||   |||
925 TACAACAGAAGGTGTTTGGGAGGC..... 949
   ||   |||   |||   |||   |||   |||   |||   |||   |||
179 erProThrThrGlnIleThrHisArgProArgValSerArg..... 192
   ||   |||   |||   |||   |||   |||   |||   |||   |||
950 ..ACCAGNAGCTCGGTGACAAACAGGCTAGAGCCACGAAGGACAGGAGG 997
   ||   |||   |||   |||   |||   |||   |||   |||   |||
193 ...AlaSerSerValAlaGlyAspLysProArg..... 202
   ||   |||   |||   |||   |||   |||   |||   |||   |||
998 GAGATGACAACTACCAAGGCTGATAGGCCAAGGGAGGCATAGAGGGGT 1047
   ||   |||   |||   |||   |||   |||   |||   |||   |||
202 ..... 202
   ||   |||   |||   |||   |||   |||   |||   |||   |||
1048 CAGGATAGCTCTTGATGCAGCCAAAGAGTCCTAGGAACCATTTGGGCCAC 1097
   ||   |||   |||   |||   |||   |||   |||   |||   |||
203 .....ThrPheLeuProSerThrThr 209
   ||   |||   |||   |||   |||   |||   |||   |||   |||
1098 CAGCTCTGGTCTCAGAACATTGGCCTGGGAATCTCCACAGCAACG 1147
   ||   |||   |||   |||   |||   |||   |||   |||   |||
210 AlaSerLysIleSerAlaLeuGluGlyLeuLeuLysProGlnThrProSe 226
   ||   |||   |||   |||   |||   |||   |||   |||   |||
1148 CCAGTTTCTAAGCAACAATCTCAGGGTTCATTGGAGAAACAACACTCCAGC 1197
   ||   |||   |||   |||   |||   |||   |||   |||   |||
226 rTyrAsnHisHisThrArg..... 232

```

```

1198 TGCAGGATGTCGACCTTGGAAACTCCAGTCGAGATGTGTGGATCTGG 1247
   :   |||   |||   |||   |||   |||   |||   |||   |||
233 .....LeuHisArgGlnArgAlaLeuAspTyr 241
   ||   |||   |||   |||   |||   |||   |||   |||   |||
1248 GAACTCCAGCTGCAGATGTGTGACCAGCATGGAGGCATCTGGGAA 1297
   ||   |||   |||   |||   |||   |||   |||   |||   |||
242 GlySerGlnSerGly..... 246
   ||   |||   |||   |||   |||   |||   |||   |||   |||
1298 GGAAGCGCTCGAGGGACCTAGATGCTGCCACTGGAGACAGAGTCCCA 1347
   ||   |||   |||   |||   |||   |||   |||   |||   |||
246 ..... 246
   ||   |||   |||   |||   |||   |||   |||   |||   |||
1348 AGCAACACTGAGCCACACCCGGCAGTAGACCTGGGACCCCTGGCA 1397
   ||   |||   |||   |||   |||   |||   |||   |||   |||
247 .....ArgGluGlyGlnGlyPheHis 253
   ||   |||   |||   |||   |||   |||   |||   |||   |||
1398 AGGAGTCTCCGTGAAGCGTACTTTTCCAGAAAGATGAAGCAGCTCGG 1447
   ||   |||   |||   |||   |||   |||   |||   |||   |||
254 IleLeuIlePro.....ThrIleLeuGlyLeuPheLeuAlaLeuLe 268
   ||   |||   |||   |||   |||   |||   |||   |||   |||
1448 ACCCTGGCTCCTGTCTCTACCATGCTGGCCCTGTTTATGCTTATGGCT 1497
   ||   |||   |||   |||   |||   |||   |||   |||   |||
268 uGlyLeuValValLysArgAlaValGluArgArgLysAlaLeuSerArgA 285
   ||   |||   |||   |||   |||   |||   |||   |||   |||
1498 GGTCTCTATTG.....CAAAGGAAGCTCTGGAGAAGA 1529
   ||   |||   |||   |||   |||   |||   |||   |||   |||
285 rgAlaArgArgLeuAlaValArgMetArgAlaLeuGluSerSerGln... 300
   ||   |||   |||   |||   |||   |||   |||   |||   |||
1580 CTGGAAGTGAACCCCAAGCAGACAGCCAGTGCCTCATGTGGAAGAAGA 1629
   ||   |||   |||   |||   |||   |||   |||   |||   |||
314 eTyr.....SerAlaCysProArgArgAlaArgGlyAlaAspAlaAlaG 329
   ||   |||   |||   |||   |||   |||   |||   |||   |||
1630 GCTCCAGGATGACTCTCTCTCT.....GCTGGGGCCAGGC 1664
   ||   |||   |||   |||   |||   |||   |||   |||   |||
329 lYThrGlyGluAlaProValProGlyPro.....Gly 339
   ||   |||   |||   |||   |||   |||   |||   |||   |||
1665 TGACTGCCCCAGAGAGAAATCCAGGACCTGAGGAGCAGAGATGAACT 1714
   ||   |||   |||   |||   |||   |||   |||   |||   |||
340 AlaProLeuPro.....ProAl 345
   ||   |||   |||   |||   |||   |||   |||   |||   |||
1715 GTCATGTTACCATGGAGAAGCAACCAAGATCAAGGCCTTCAGGACCCA 1764
   ||   |||   |||   |||   |||   |||   |||   |||   |||
345 aProLeuGlnValSerGluSerPro 353
   ||   |||   |||   |||   |||   |||   |||   |||   |||
1765 GCCTCTTTCATCATCTCTCTCCA 1789
   ||   |||   |||   |||   |||   |||   |||   |||   |||
seq_name: N_Geneseq_36:T31291
seq_documentation_block:
ID T31291 standard; cDNA; 3269 BP.
AC T31291;
DT 24-FEB-1997 (first entry)
KW Rat poly-immunoglobulin receptor, cDNA.
KW heavy chain; antigen binding domain; protection protein; mutants;
KW mucosal; environment; gastrointestinal; passive; immunisation;
KW Guy's 13 antibody; prevention; dental caries; Streptococcus;
KW poly; sorbinus; ss.
OS Rattus rattus.
FH Key Location/Qualifiers
FT cds 74..2383
FT /tag= a
PN WO9621012-A1.
PD 11-JUL-1996.
PF 27-DEC-1995; U16889.
PR 30-DEC-1994; US-367395.
PR 04-MAY-1995; US-434000.
PA (PLAN-) PLANT BIOTECHNOLOGY INC.

```





```

74  yArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluV 91
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
255 CAGGCTACCTCACCACCTCCCGGAGACGGCACATTCGTGGTGAACA 304
91  alThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGly 107
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
305 TTGCCAGCTGAGCCAGGATGACTCCGGCGCTCAAGTGTGGCTGGC 354
108 MetAsnThr..... 110
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
355 ATCAATAGCCGAGGCTGCTCCTTTGATGTCACGCTGGAGGTACGCCAGG 404
111 ..... 115
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
405 TCCTGGCTCCTAATGACACTAAAGTCTACACAGTGGACCTGGCGAGAA 454
115 hrGlnLysValThrLeuAsnValHisSerGluTyrGluProSerTrpGlu 131
|| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
455 CG.....GTGACCATCAACTGCCCTTTCAAGACTGAGATGCTCAAAAG 498
132 GluGln.ProMetProGlu.....ThrProLysTrpPheHisL 144
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
499 AGGAAGTCTTGTACAACGACAGATGGCTGTACCTGTGCTGTCATCGA 548
144 euProTyrLeuPheGlnMetProAlaTyrAlaSerSerLysPheVal 160
|| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
549 CTCACAGTGTATGTAATCCCACTATACAGGAAGATAGCCCTTGATA 598
161 ThrArgVal..... 163
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
599 TTCAGGGTACTGCCAGTACTGTTCAGCGTGTGTCATCAACCAACTCAGG 648
164 ..... 175
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
649 CTCACGATGCTGGGCACTATCTGCGAGCGCTGGGATGATTCCAATAG 698
175 alHisHisSerSerProThrThrGlnIleThrHisArgProArgValSer 191
|| |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
699 TAATAAGAAGAATGCTGACC..... 718
192 ArgAlaSerSerValAlaGlyAspLysProArgThrPheLeuProSerTh 208
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
719 .....TCCAAAGTCTAAAGCCGACCGCGAGCTGGTTATGAAGACCTC 762
208 rThrAlaSerLysIleSerAlaLeuGluGlyLeuLeuLysProGlnThr. 224
; |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
763 AGGGCTCAGTGACCTCCACTGTGCCCTGGGCCCTGAGGTGGCAACGT 812
225 ..ProSerTyrAsnHisHisThrArgLeuHisArg..... 235
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
813 GGCCAAATTTCTGTCCCGACAGACAGTCGGGAAACCTGTGCGTGC 862
236 .....GlnArgAlaLeuAspTyrGlySerGlnSe 245
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
863 TCAACACCTTAGGAGAGAGGCCGCCAGCTTTGAGGCGCAGGATCCTGCT 912
245 r..... 246
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
913 AACCCCGAGGACAGGATGGCTCATTCAGTGTGCTGATCACAGCCGTGAG 962
246 lyArgGluGlyGlnGlyPheHisIleLeuIleProThrIleLeuGlyLeu 262
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
963 GAAGGAGGATGACGGCGCTACCTGTGTGGAGCCATT..... 1000
263 PheLeuLeuAlaLeuLeuGlyLeuValValLysArgAlaValGluArgAr 279
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1001 .....CGGATGCTCAGCTGCAG 1017
279 gLysAlaLeuSerArgArgAlaArgLeuAlaValArgMetArgAlaL 296
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1018 GAAGGCT.....CGCCTATCCAGGC 1037

```

```

296 euGluSerSerGlnArgProArgGlySerProArgProArgSer..... 310
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1038 CTGGCAACTCTTCGTCATGAGAGTCCACCATTCCTCCCGCAGCCCATCG 1087
311 .....GlnAsnAsnIleTyr...SerAlaCysProArgAr 321
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1088 TGTGTAAGGGGTGGCAGGAGCTCTGTGGCGTCTGCTGCCCCCTACAAAC 1137
321 gAlaArgGlyAlaAspAlaAlaGlyThrGlyGluAlaProValProGlyP 338
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1138 CGTAAGGAAAGCAAAAGCATCAAGTACTGTT.....GTCTCTGGGA 1178
338 roGlyAlaProLeuProAlaPro..... 346
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1179 AGGGCCCAAGATGGCGCTGCCCTCTGCTGTGGACAGCGAGGGGTGGG 1228
347 .....LeuGlnValSerGluSerProTyr.....LeuHi 356
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1229 TTAAGCCCACTAGCAGGCGCGCTCTCCTCTGCTGGAGGAGCCAGGCAC 1278
356 salaProSerLeuLysThrSer 363
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1279 GGCACCTTCACCTGTCATCTCA 1300
seq_name: N_Geneseq_36:X07407
seq_documentation_block:
ID X07407 standard; cdna to mRNA; 1839 BP.
AC X07407;
DT 08-JUN-1999 (first entry)
DE Human secretory immunoglobulin A component gene.
KW Immunoglobulin A; secretory; component; IGA; human; treatment;
KW prevention; infection; HIV; AIDS; cold; flu; virus; gene;
KW human immunodeficiency virus; respiratory syncytial virus; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 1..1839
FT /tag= a
FT /product= secretory component
FN WO9857993-A1.
PD 23-DEC-1998.
PF 10-JUN-1998; U11975.
PR 19-JUN-1997; US-050969.
PA (REGC ) UNIV CALIFORNIA.
PI Chintalacharuvu KR, Morrison SL;
DR WPI; 99-080950/07.
DR P-PSDB; W95601.
PT Producing secretory immunoglobulin in single cells - useful to
PT produce commercial quantities of secretory immunoglobulin to prevent
PT or treat infections
PS Disclosure; Page 22; 39pp; English.
CC The sequence is that of the coding region for the secretory
CC component of human secretory immunoglobulin A (siga).
CC The sequence can be used as part of a method for the
CC production of sig molecules. This method is useful for
CC producing commercial quantities of sig (especially siga) to treat
CC or prevent infections. In particular, siga produced by the method
CC can be used to prevent or treat infections in mammals, birds or
CC fish; especially systemic infections or infections at a mucosal
CC surface. It is especially useful to prevent or treat infection
CC with human immunodeficiency virus (HIV), respiratory syncytial
CC virus, flu virus or cold virus. The method allows production of
CC commercial quantities of sig molecules for therapeutic use, not
CC previously possible; production using non-plant cells and a
CC single cell type is more efficient than a previous multi-step
CC process of fusing recombinant plant cells, and avoids alterations
CC of the sig by plant cells. SIGA molecules are more stable
CC and resistant to proteolysis than previously used IGA molecules,
CC and can be administered to prevent as well as to treat infections,
CC unlike e.g. IgG and IgM molecules.
SQ Sequence 1839 BP; 423 A; 504 C; 544 G; 368 T;

```

```

alignment_scores:
  Quality: 182.00      Length: 115
  Ratio: 2.247        Gaps: 5
  Percent Similarity: 70.435      Percent Identity: 37.391

alignment_block:
  US-09-135-238B-2 x X07407 ..
  Align seg 1/1 to: X07407 from: 1 to: 1839

12 ProValSerGlyAlaLeuArgIleLeuProGluValLysValGluGlyL 28
AC T30856;
DT 13-SEP-1996 (first entry)
DE Secretory component DNA fragment 1.
KW Secretory component; SC DNA fragment 1; crystallisation;
OS Polymeric immunoglobulin receptor; pIgR; CHO; ss.
OS Chimeric Homo sapiens;
OS Chimeric synthetic.
FH Key Location/Qualifiers
FT misc_feature 1..33
FT /tag= a
FT /note= "bases 1-33 derived from PCR with 5'
FT primer 1"
FT 111..2015
FT /tag= b
FT /product= transcript from SC DNA fragment 1
FT misc_feature complement (1997..2031)
FT /tag= c
FT /note= "from PCR with 3' primer"
PD WO9618734-A1.
PN 20-JUN-1996.
PR 06-DEC-1995; E04797.
PR 16-DEC-1994; EP-120019.
PA (CIBA ) CIBA GEIGY AG.
PI Asselbergs FAM, Gruetter M, Leist C, Van Oostrum J;
DR WPI: 96-300651/30.
DE Prodn. of recombinant secretory protein capable of crystallisation -
PT Pref. human poly-IgR, for use in screening and binding studies
PS Claim 5; Page 36-38, 50pp; English.
CC SC DNA fragment 1 (T30856) was obtd. by PCR amplification of human
CC polymeric immunoglobulin receptor (pIgR) cDNA in vector pCB6 using a
CC 5' primer (T30859) hybridising to the cytomegalovirus promoter in

```

```

CC pCB6 and a 3' primer (T30861) that generates a stop codon at
CC position 1906 just upstream of the hydrophobic transmembrane segment
CC of the pIgR and also creates a downstream XbaI site. CHO SSF 3
CC cells transfected with a vector carrying the amplified DNA can be
CC used for prodn. of recombinant secretory component (SC). This can
CC be crystallised for receptor structure studies, used to stabilise
CC immunoglobulins, or used to screen (ant)agonists capable of
CC modulating mucosal immune responses.
SQ Sequence 2031 BP; 477 A; 541 C; 605 G; 408 T;

alignment_scores:
  Quality: 182.00      Length: 115
  Ratio: 2.247        Gaps: 5
  Percent Similarity: 70.435      Percent Identity: 37.391

alignment_block:
  US-09-135-238B-2 x T30856 ..
  Align seg 1/1 to: T30856 from: 1 to: 2031

12 ProValSerGlyAlaLeuArgIleLeuProGluValLysValGluGlyL 28
AC T30856;
DT 24-FEB-1997 (first entry)
DE Human poly-immunoglobulin receptor, cDNA.
KW Human; immunoglobulin; receptor; protection protein; mutants;
KW heavy chain; antigen binding domain; protection; pathogen;
KW mucosal; environment; gastrointestinal; passive; immunisation;
KW Guy's 13 antibody; prevention; dental caries; Streptococcus;
KW poly; sorbinus; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 235..2475
FT /tag= a
FT WO9621012-A1.
PN 11-JUL-1996.
PR 27-DEC-1995; U16889.
PR 30-DEC-1994; US-367395.
PR 04-MAY-1995; US-434000.
PA (PLAN-) PLANT BIOTECHNOLOGY INC.
PA (UNWE-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.
PA (PLAN-) PLANET BIOTECHNOLOGY INC.

```

```
seq_name: N_Geneseq_36:T30856
```

```
seq_documentation_block:
```

```
ID T30856 standard; cDNA; 2031 BP.
```

```
AC T30856;
```

```
DT 13-SEP-1996 (first entry)
```

```
DE Secretory component DNA fragment 1.
```

```
KW Secretory component; SC DNA fragment 1; crystallisation;
```

```
OS Polymeric immunoglobulin receptor; pIgR; CHO; ss.
```

```
OS Chimeric Homo sapiens;
```

```
OS Chimeric synthetic.
```

```
FH Key Location/Qualifiers
```

```
FT misc_feature 1..33
```

```
FT /tag= a
```

```
FT /note= "bases 1-33 derived from PCR with 5'
```

```
FT primer 1"
```

```
FT 111..2015
```

```
FT /tag= b
```

```
FT /product= transcript from SC DNA fragment 1
```

```
FT misc_feature complement (1997..2031)
```

```
FT /tag= c
```

```
FT /note= "from PCR with 3' primer"
```

```
PD WO9618734-A1.
```

```
PN 20-JUN-1996.
```

```
PR 06-DEC-1995; E04797.
```

```
PR 16-DEC-1994; EP-120019.
```

```
PA (CIBA ) CIBA GEIGY AG.
```

```
PI Asselbergs FAM, Gruetter M, Leist C, Van Oostrum J;
```

```
DR WPI: 96-300651/30.
```

```
DE Prodn. of recombinant secretory protein capable of crystallisation -
```

```
PT Pref. human poly-IgR, for use in screening and binding studies
```

```
PS Claim 5; Page 36-38, 50pp; English.
```

```
CC SC DNA fragment 1 (T30856) was obtd. by PCR amplification of human
```

```
CC polymeric immunoglobulin receptor (pIgR) cDNA in vector pCB6 using a
```

```
CC 5' primer (T30859) hybridising to the cytomegalovirus promoter in
```

```
seq_name: N_Geneseq_36:T31288
```

```
seq_documentation_block:
```

```
ID T31288 standard; cDNA; 2919 BP.
```

```
AC T31288;
```

```
DT 24-FEB-1997 (first entry)
```

```
DE Human poly-immunoglobulin receptor, cDNA.
```

```
KW Human; immunoglobulin; receptor; protection protein; mutants;
```

```
KW heavy chain; antigen binding domain; protection; pathogen;
```

```
KW mucosal; environment; gastrointestinal; passive; immunisation;

```

```
KW Guy's 13 antibody; prevention; dental caries; Streptococcus;

```

```
KW poly; sorbinus; ss.
```

```
OS Homo sapiens.
```

```
FH Key Location/Qualifiers
```

```
FT cds 235..2475
```

```
FT /tag= a
```

```
FT WO9621012-A1.
```

```
PN 11-JUL-1996.
```

```
PR 27-DEC-1995; U16889.
```

```
PR 30-DEC-1994; US-367395.
```

```
PR 04-MAY-1995; US-434000.
```

```
PA (PLAN-) PLANT BIOTECHNOLOGY INC.
```

```
PA (UNWE-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.
```

```
PA (PLAN-) PLANET BIOTECHNOLOGY INC.
```



119 rLeuAsnValHisSerGlu 125  
:||||:||||| :|||:  
523 CCTGGAGTCAGCAGAT 541

seq\_name: N\_Geneseq\_36:T96033

seq\_documentation\_block:  
ID T96033 standard; cDNA; 2084 BP.

AC T96033;  
DE 21-MAY-1998 (first entry)  
DE Rat kidney injury related molecule (KIM) cDNA clone 1-7.  
KW Kidney injury related molecule; KIM; rat; renal disease; injury;  
KW nephritis; tissue regeneration; therapy; ss.  
OS Rattus sp.  
FH Key Location/Qualifiers  
FT CDS 145..1068  
FT CD5 /\*tag= a  
FN W09744460-A1.  
PD 27-NOV-1997.  
PR 23-MAY-1997; U09303.  
PR 23-AUG-1996; US-023442.  
PR 24-MAY-1996; US-018228.  
PA (BIOJ ) BIOGEN INC.  
PI Bonventre JV, cate RL, Hession CA, Ichimura T, Sanicola-Nadel M,  
PI Wei H;  
DR WPI; 98-018514/02.  
DR P-PSDB; W38334.  
PT DNA encoding kidney injury related molecule - which is upregulated  
PT in injured or regenerating tissue. useful to promote growth of new  
PT tissue and survival of damaged tissue  
PS Claim 1; Page 34-37; 68pp; English.  
CC cDNA clone 1-7, deposited as ATCC 98060, codes for a rat protein  
CC (see W38334), designated kidney injury related molecule (KIM), that  
CC is up-regulated in injured or regenerating tissue. Representational  
CC difference analysis was used to examine cDNA libraries prepared from  
CC ischaemic and normal rat adult kidneys. 3 Cycles of hybridisation  
CC and selective amplification gave 3 fragments present only in the  
CC injured kidney library. Screening of the injured kidney library  
CC with these fragments gave clone 1-7, as well as clone 3-2 (see  
CC T96033), which is a splice variant of 1-7 encoding an identical KIM  
CC protein (nucleotides 136-605 of 3-2 represent an insertion), and  
CC clone 4-7 (see T96034) encoding a 307-amino acid KIM protein (see  
CC W38335). A labelled probe from clone 1-7 was used to identify a  
CC human KIM sequence (see T96035). KIM nucleic acids can be used in  
CC the recombinant production of KIM polypeptides in prokaryotic or  
CC eukaryotic host cells. KIM, or an agonist, can be used to treat renal  
CC disease and to promote the growth of new tissue or the survival of  
CC damaged tissue, generally in conditions where the binding of  
CC specific ligand to KIM stimulates cell growth, maintains cellular  
CC differentiation or reduces apoptosis, e.g. in cases of renal  
CC failure, nephritis, kidney transplants, toxic or hypoxic injury.  
CC Damage/regeneration of renal cells can be determined by measuring  
CC KIM polypeptide or nucleic acid, e.g. with an antisense probe,  
CC particularly to diagnose or monitor the progress of disease or  
CC therapy.

Sequence 2084 BP; 604 A; 451 C; 470 G; 559 T;  
alignment\_scores:  
Quality: 171.00 Length: 403  
Ratio: 0.924 Gaps: 22  
Percent Similarity: 45.906 Percent Identity: 24.318  
alignment\_block:  
US-09-135-238B-2 x T96033 ..  
Align seg 1/1 to: T96033 from: 1 to: 2084

13 ValSerGlyAlaLeuArgIleLeuPro.....GluVa 23  
:||||| :|||:  
166 ATTTCAGGCTCTGCTGCTTCTCCAGGCTCTGTAGATTCATTATGAAGT 215  
23 llysValGluGlyLeuGlySerValThrIleLysCysProLeup 40

! |||:||||| :|||:  
216 A...GTGAAGGGGTGGTGGGTCAACCTGTGCACAAATTCATGTAATCTACT 262  
40 roGluMetHisValArgIleTyrLeuCys.....ArgGluMetAlaGly 54  
||| :|||:  
263 CAACACGTGGAGGAATCAACACGACATGTTGGGGCCGGGGCAATGCCCA 312  
55 SerGlyThrCysGlyThrValValSerThrThrAsnPhelLysAlaG 71  
:||||| :|||:  
313 TATTCTAGTTGTTCAATATATCTATTGGACCAATGATACCAAGTAC 362  
71 uTyrLys.....GlyArgValThrLeuLysGlnTyrProArgLysAsn 86  
:||||| :|||:  
363 CTATCGGAGCAGCGTGTGATACACATATAAGGGCGGTATTTCAGAAGGAG 412  
86 euPheLeuValGluValThrGlnLeuThrGluSerAspSerGlyValTyr 102  
:||| :|||:  
413 ACGTATCTCTGACAAATAGAGAACTCTGTGTATGATGATGATGCTGTAT 462  
103 AlaCysGlyAla.....GlyMetAsnThrAspArgGlyLysThrG 116  
:||||| :|||:  
463 TGTTCGGAGTGGAGATTCTCTGGATGGTTCAACGAT.....CA 500  
116 nLysValThrLeuAsnValHisSerGluTyrGluProSerTyrGluGlu 133  
:||||| :|||:  
501 GAAATGACCTTTTCATTG.....GAAGTTAAACCAAGAAATCCACAA 544  
133 InProMetProGluThrProLysTyrPheHisLeuProTyrLeuPheGln 149  
:||||| :|||:  
545 GTCT...CCAACAAGACCC..... 561  
150 MetProAlaTyrAlaSerSerSerLysPheValThrArgValThrPr 166  
:||||| :|||:  
562 .....ACAACTACAGACCCACACCAAGGCCCAACTAT 599  
166 oAlaGlnArgGlyLysValProPro.....ValHisHisSers 179  
:||||| :|||:  
600 TTCAACAGATCCACATGTACCAACATCAACAGAGTCTCCACTCTA 649  
179 erProThr.....ThrGlnIleThrHisArgProArgValSerArg 192  
:||||| :|||:  
650 CTCACACACGACCAACACAG...ACTCACAACCAAGAAATCACT... 693  
193 AlaSerSerValAlaGlyAspLysProArgThrPheLeuProSerThrTh 209  
:||||| :|||:  
694 .....ACATTTCATGCCCATGAGAC 713  
209 rAlaSerLysIleSerAlaLeuGluGlyLeuLysProGlnThrPro 226  
:||||| :|||:  
714 AACTGTGAGGTGACA.....GAACTCCAT 739  
226 erTyr..... 227  
740 CATATATCTCTGACACTGGAATGGCACTGTGCATCTCTCAGAGAGGCC 789  
228 .....AsnHisThrArgLeuHisArgGlnArgAlaLeuAspTyrG 242  
:||||| :|||:  
790 TGGNAATATCACACTGTAAAGATCCCTTTGAGGAAG..... 825  
242 ySerGlnSerGlyArgGluGlnGlyPheHisIleLeuIleProThrI 259  
:||||| :|||:  
826 ...CCGCAGAGAAACCGACTAAGGGCTTCTATGTTGGCATGTCGGTTG 871  
259 leLeuGlyLeuPheLeuAlaLeuLeuGlyLeuValValLysArgAla 275  
:||||| :|||:  
872 CAGCCCTGCTGCTGCTGCTGCTTCCGAGCACCGTGGTGTCCACAGGTAC 921  
276 ValGluArgGlys.....AlaLeuSerArgArgAlaArgLe 289  
:||||| :|||:  
922 ATCATATTAAGAAAGAGATGGGCTCTCTGAGCTTTGTTGCCTTCATGT 971  
289 uAlaValArgMetArgAlaLeuGluSerSerGlnArgProArgGlySerP 306  
:||||| :|||:

```
972 CTCT...AAGAGTAGAGCTTTGCAGACGCA.....GGCATTG 1006
306 roArgProArgSerGlnAsnAsnIleTyrSerAlaCysProArgArgAla 322
   ::::::::::::::::::::::::::::|
1007 TCATCCCGAGCTGAAGACAACACTAC...ATTATTGAAGATAGATCT 1053
   ::::::::::::::::::::|
323 ArgGlyAlaAsp..... 326
   ::::::::::::::|
1054 CGAGGTGCAGATAGTCCAGAGGCTTCGTGGGGCTTCTGCTCGG 1103
327 .....AlaAlaGlyThrGlyGluAlaProValProGlyP 338
   ::::|
1104 ATTACAGAGATCGTGACTGATTTCACAGAGTAAATACCCATTCCAGCTC 1153
338 roGly.....AlaProLeuProProAlaProLeuGlnValSer 350
   ::::|
1154 CTGGGAGATTTGTGTTTGGTTCCTCCAGCTGCAGTGGAGAGGTAACC 1203
351 GluSerPro 353
   |||
1204 CTCTACCCT 1212
```

seq\_name: N\_Geneseq\_36:T96032

seq\_documentation\_block:

AC T96032 standard; cDNA; 2566 BP.

DT 21-MAY-1998 (first entry)

KW Rat kidney injury related molecule (KIM) cDNA clone 3-2.

OS Kidney injury related molecule; KIM; rat; renal disease; injury; nephritis; tissue regeneration; therapy; ss.

OW Rattus sp.

EH Key Location/Qualifiers

FT CDS 615..1538

ET /\*tag= a

FN WO9744460-A1.

PD 27-NOV-1997.

PF 23-MAY-1997; U09303.

PR 23-AUG-1996; US-023442.

PR 24-MAY-1996; US-018228.

FA (BIOJ) BIOGEN INC.

PI Bonventre JV, Cate RL, Hession CA, Ichimura T, Sanicola-Nadel M,

DR Wei H;

DR WPI; 98-018514/02.

DR P-PSDB; W38334.

PT DNA encoding kidney injury related molecule - which is upregulated

PT in injured or regenerating tissue, useful to promote growth of new

PT tissue and survival of damaged tissue

PS Claim 1; Page 32-34; 68pp; English.

CC cDNA clone 3-2, deposited as ATCC 98061, codes for a rat protein

CC (see W38334), designated kidney injury related molecule (KIM), that

CC is up-regulated in injured or regenerating tissue. Representational

CC difference analysis was used to examine cDNA libraries prepared from

CC ischaemic and normal rat adult kidneys. 3 Cycles of hybridisation

CC and selective amplification gave 3 fragments present only in the

CC injured kidney library. Screening of the injured kidney library

CC with these fragments gave clone 3-2, as well as clone 1-7 (see

CC T96033), which is a splice variant of 3-2 encoding an identical KIM

CC polypeptide (nucleotides 136-605 of 3-2 represent an insertion),

CC and clone 4-7 (see T96034) encoding a 307-amino acid KIM protein

CC (see W38335). Clone 1-7 was used to identify a human KIM sequence

CC (see T96035). KIM nucleic acids can be used in the recombinant

CC production of KIM polypeptides in prokaryotic or eukaryotic host

CC cells. KIM, or an agonist, can be used to treat renal disease and to

CC promote the growth of new tissue or the survival of damaged tissue,

CC generally in conditions where the binding of specific ligand to KIM

CC stimulates cell growth, maintains cellular differentiation or

CC reduces apoptosis, e.g. in cases of renal failure, nephritis,

CC kidney transplants, toxic or hypoxic injury. Damage/regeneration

CC of renal cells can be determined by measuring KIM polypeptide or

CC nucleic acid, e.g. with an antisense probe, particularly to

CC diagnose or monitor the progress of disease or therapy.

CC sequence 2566 BP; 726 A; 546 C; 591 G; 703 T;

SQ

```
alignment_scores:
  Quality: 171.00      Length: 403
  Ratio: 0.924        Gaps: 22
  Percent Similarity: 45.906      Percent Identity: 24.318

alignment_block:
US-09-135-238B-2 x T96032 ..
Align seg 1/1 to: T96032 from: 1 to: 2566
13 ValSerGlyAlaLeuArgIleLeuPro.....GluVa 23
   ::::::::::| | | ::::::::::|
636 ATTTGAGGCTCTGCTGCTTCTCCAGGCTGTGTAGATCTTATGAAGT 685
23 lLysValGluGluLeuGlySerValThrIleLysCysProLeuP 40
   | | | | | | | | | | | | | | | | |
686 A...GTGAAGGGGTGTGGGTCACTGTGCAAAATCCATGTACTTACT 732
40 roGluMetHisValArgIleTyrLeuCys.....ArgGluMetAlaGly 54
   ||| ||| |||
733 CAACAGCTGGAGGAATCAACAACGACATGTTGGGGCCGGGGCAATGCCCA 782
55 SerGlyThrCysGlyThrValValSerThrThrAsnPheIleLysAlaGl 71
   ::::::::::| | | | | | | | | | | | | | |
783 TATTTAGTTGTTCAAAATATACTTATTTGGACCAATGGATACCAAGTCAC 832
71 uTyrLys.....GlyArgValThrLeuLysGlnTyrProArgLysAsnL 86
   || | | | | | | | | | | | | | | | | |
833 CTATCGGAGCAGCGGTGCGATACAACAATAAAGGGCGTATTTCAGAAGGAG 882
86 euPheLeuValGluValThrGlnLeuThrGluSerAspSerGlyValTyr 102
   ::::| | | | | | | | | | | | | | | | |
883 ACGTATCTCTGACAATAGAGAACTCTGTTGATAGTAGTAGTGTGTGTAT 932
103 AlaCysGlyAla.....GlyMetAsnThrAspArgGlyLysThrGl 116
   ::::| | | | | | | | | | | | | | | | |
933 TGTTCGGAGTGGAGATTCTGTTGATGTTCAACGAT.....CA 970
116 nLysValThrLeuAsnValHisSerGluTyrGluProSerTrpGluGluG 133
   || | | | | | | | | | | | | | | | |
971 GAAATGACCTTTTCATTG.....GAAGTTAAACCAGAAATTCACCAA 1014
133 lnProMetProGluThrProLysTrpPheHisLeuProTyrLeuPheGln 149
   || | | | | | | | | | | | | | | | |
1015 GTCTT...CCAAACAAGACCC..... 1031
150 MetProAlaTyrAlaSerSerLysPheValThrArgValThrThrPr 166
1032 .....ACAACTACAGACCCCAACCAAGGCCCAACTAT 1069
166 oAlaGlnArgGlyLysValProPro.....ValHisHisSers 179
   :: | | | | | | | | | | | | | | | | |
1070 TTCAACAAGATCCACACATGTACCAACATCAACAGAGTCTCCACCTTA 1119
179 erProThr.....ThrGlnIleThrHisArgProArgValSerArg 192
   || | | | | | | | | | | | | | | | |
1120 CTCCAACACCAAGAACAAACACAG...ACTCAAAACACAGAAATCACT... 1163
193 AlaSerSerValAlaGlyAspLysProArgThrPheLeuProSerThrTh 209
1164 .....ACATTTTATGCCCATGAGAC 1183
209 rAlaSerLysIleSerAlaLeuGluGlyLeuLeuLysProGlnThrProS 226
   | | | | | | | | | | | | | | | | |
1184 AACTGCTGAGTGACA.....GAAACTCCAT 1209
226 erTyr..... 227
   || | | |
1210 CATATCTCTCTGAGACTGGAATGGCACTGTGACATCTCTCAGAGGAGGCC 1259
228 .....AsnHisThrArgLeuHisArgGlnArgAlaLeuAspTyrGl 242
```

1260 TGGATAATACACTGTAGAAATCCCTTCAGGAG... 1295  
 242 yserGlnSerGlyArgGluGlyGlnGlyLeuIleProThrI 259  
 1296 ...CCGCAGAGAAACCGCACTAGGCGTTCATATGGCATGCGGTG 1341  
 259 leuLeuClyLeuPheLeuLeuAlaLeuLeuClyLeuValValValValArgAla 275  
 1342 CAGCCCTCGTCTGCTGCTGCTTGGAGCACCGTGGTGTGCACAGGTAC 1391  
 276 ValGluArgArgLys.....AlaLeuSerArgArgAlaArgArgLe 289  
 1392 ATCATTTATGAAGAAGATGGGCTCTCTGAGCTTTGTGSCCTTCGATGT 1441  
 289 uAlaValArgMetArgAlaLeuGluSerSerGlnArgProArgGlySerP 306  
 1442 CTCT..AAGATAGAGCTTTGCAGAACGCA.....GCGATTG 1476  
 306 roArgProArgSerGlnAsnAsnIleTyrSerAlaCysProArgArgAla 322  
 1477 TGCATCCCGAGCTGGAAGACAACATCTAC...ATTATTGAAGATAGATCT 1523  
 323 ArgGlyAlaAasp..... 326  
 1524 CGAGGTGCAGAATGAGTCCAGAGSCCTTCTGTGGGCGCTTCTGCCTGGG 1573  
 327 .....AlaAlaGlyThrGlyGluAlaProValProGlyP 338  
 1574 ATTACAGATCGTGACTGATTTCCAGAGATAAAATACCCCATTCAGAGTC 1623  
 338 roGly.....AlaProLeuProAlaProLeuGlnValSer 350  
 1624 CTGGAGATTTTGTGTTTGGTTCTTCCAGCTCGAGTCGAGAGGGTAAAC 1673  
 351 GluSerPro 353  
 1674 CTCTACGCT 1682

seq name: N Geneseq 36:014498

```

seq_documentation_block:
ID      Q14498 standard: DNA; 1876 BP.
AC      Q14498:
AD      10-MAR-1993 (revised)
DT      30-JAN-1992 (DT
DE      Clone 1 for truncated poly Ig-receptor.
KW      Rabbit; insemination; pregnancy; ss.
OS      Oryctolagus cuniculus.
FH      Key      Location/Qualifiers
          cds      124..1876
          FT      W09116061-A.
          PD      31-OCT-1991.
          PF      16-APR-1991; U02604.
          PR      16-APR-1990; US-510161.
          PA      (HARD ) HARVARD COLLEGE.
          PA      (SURE-) INST SUISSE RECH EXPER C.
          PI      Krachenbuhl JP, Weltzin RA, Neutra MR;
          WP1: 91-339549/46.
          DR      P-PSDB; R14670.
          DR      Stabilised poly-Ig complex contg. portion of poly-Ig receptor -
          PT      useful in protection against pathogens or against pregnancy
          PS      Disclosure; Fig 8; 51 pp; English.
          CC      The sequence was obt'd. from a cDNA.
          CC      encodes a truncated poly-Ig receptor. The native gene (Mostov et
          CC      al.) is mutated to delete the portion encoding the transmembrane and
          CC      intra-cellular domains. The recombinant protein produced by
          CC      expression of the sequence is used as a stabiliser protein with a
          CC      poly-Ig specific for a selected antigen or family of antigens. The
          CC      compsn. can be administered directly to the mucosal surfaces of a
          CC      mammal to protect against a pathogen or against insemination. It
          CC      protects against allergens that contact the respiratory or digestive
          CC      mucosal surfaces and protects against pregnancy by cross-linking
          CC

```

```

CC sperm in the vagina.
CC See also Q14498;
CC Sequence 1876 BP; 441 A; 525 C; 549 G; 361 T;

alignment_scores:
    Quality: 166.50      Length: 103
    Ratio: 2.602         Gaps: 4
    Percent Similarity: 62.136      Percent Identity: 42.718

alignment_block:
US-09-135-238B-2 x Q14498 ..

Align seg 1/1 to: Q14498 from: 1 to: 1876

30 GlyGlySerValThrIleLysCysProLeuPro.....GluMe 42
   ||| |||||:||||| ||| |||
238 GGCAGCTCGGTGTCATCATCTACTACCCACAACCTCGCTCACCG 287
   ||| |||||:||||| ||| |||
42 tHisValArgIleTyrLeuCysArgGluMetalArgSerGlyThrCysG 59
   ||||| |||||:||||| ||||| |||||
288 GCACACCGGAGTCTGCTGCCGGAA...GAGGAGCGCGCGCTGG 334
   ||||| |||||:||||| ||||| |||||
59 lyThrValValSerThrThrAsnPhelIleLysAlaGluTyrLysGlyArg 75
   ||||| |||||:||||| ||||| |||||
335 TGACGCTTGCTCG...ACCGGCTACAGTCCACAGGAATACTCCGGGAGA 381
   ||||| |||||:||||| ||||| |||||
76 ValThrLeuLysGlnTyrProAlgLysAsnLeuPheLeuValGluValTh 92
   ||||| |||||:||||| ||||| |||||
382 GGCAGCTCACCGACTCCCTCGTATAAGGGGAGTTGTGGTGACTGTGA 431
   ||||| |||||:||||| ||||| |||||
92 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMetA 109
   ||||| |||||:||||| ||||| |||||
432 CCAACTCACCCAGACAGCACTCAGGGAGCTACAAGTGTGGGTGGGAGTCA 481
   ||||| |||||:||||| ||||| |||||
109 snThrAspArgGlyThrGlnLysValThrLeuAsnValHisSerGlu 125
   ||||| |||||:||||| ||||| |||||
482 AC...GCCGCTGGCCTGGACTTCGGTGTCAACGTGCTGGTCAGCCAGAAG 528
   ||||| |||||:||||| ||||| |||||

126 TyrGluPro 128
   |||||
529 CCAGAGCCT 537

```

Tue Sep 12 08:57:54 2000

PS Claim 10; Pages 99-102; 152pp; English.  
 CC The present sequence encodes the rabbit poly-immunoglobulin (Ig)  
 CC receptor, a portion of which corresp. to residues 1-627, pref.  
 CC 1-606, or esp. residues 21-43, 1-118, 119-223, 224-332, 333-441,  
 CC 442-552, 553-606 or 553-627 comprises a protection protein (PP).  
 CC The Ig of the invention comprises a PP as above in association with  
 CC an Ig derived heavy chain, having at least a portion of an antigen  
 CC (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.  
 CC gastrointestinal, environments, therefore enhancing its  
 CC effectiveness in passively immunising animals against mucosal  
 CC pathogens. The Ag binding domain is specifically derived from the  
 CC guy's 13 antibody, and the Ig can be used to prevent dental caries  
 CC by binding, e.g. Streptococcus mutans serotypes c, e and f, or  
 CC S. sorbinus serotypes d and g.  
 SQ Sequence 3517 BP; 795 A; 1027 C; 1038 G; 657 T;

## alignment\_scores:

Quality: 166.50 Length: 103  
 Ratio: 2.602 Gaps: 4  
 Percent Similarity: 62.136 Percent Identity: 42.718

## alignment\_block:

US-09-135-238B-2 x T31287 ..

Align seg 1/1 to: T31287 from: 1 to: 3517

30 GlyGlySerValThrIleLysCysProLeuPro.....GluMe 42  
 ||| |||||:|||| ||| |||  
 238 GCGNCTCGGTGTCATCATGCTACTACCCACACACTCCGTCACCCG 287  
 42 tHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 59  
 ||| ||| :| ||||| ||| ||||| |||  
 288 GCACAGCCGGAAGTCTGCTGCCGGAA...GAGGAGAGCGCGCTGCG 334  
 59 lyThrValValSerThrThrAsnPheIleLysAlaGluTyrLysGlyArg 75  
 |||:||||| |||:||||| :|||:|||||:|||||  
 335 TGACGCTTGCCCTCG...ACCGCTACACGTCGCCAGGAATACTCCGGAGA 381  
 76 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 92  
 ||| :|||:|||| |||:| |||:| |||  
 382 GGCAAGCTCACCGACTCCCTGTATAAAGGGAGTTTGTGTGACTGTTGA 431  
 92 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 109  
 |||||:|||||:||||| ||| |||||:|||||:|  
 432 CCAACTCACCAGAACGACTCAGGGAGCTACAAGTGTGGCGTGGAGTCA 481  
 109 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 125  
 || ||||| |||||:| |||:| |||:|  
 482 AC...GGCCGTGGCTGGACTCGGTGTCACACGTGCTGCTCAGCCAGAG 528  
 126 TyrGluPro 128  
 |||||  
 529 CCAGAGCCT 537





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2000, 10:09:19 ; Search time 88.29 Seconds  
(without alignments)  
5412.469 Million cell updates/sec

Title: US-09-135-238B-1

Sequence: 1 aaaggaagcagcagcgtgc.....ttactctgtccatcctt 1910

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1830.4	95.8	2040	1 X28178	Human PIGRL-1 codi
2	676	35.4	1047	1 X28179	Human PIGRL-1 codi
3	45.2	2.4	117213	1 V62176	HSV-2 strain SB5 C
4	41	2.1	1911	1 V20383	cDNA for human imm
5	40.4	2.1	1984	1 V19114	Human secreted apo
6	40.2	2.1	987	1 T89157	Oerskovia xanthine
7	40.2	2.1	1516	1 T89156	Oerskovia xanthine
8	39.6	2.1	3850	1 X21656	Human helicase gen
9	39	2.0	117213	1 V62176	HSV-2 strain SB5 C
10	38.6	2.0	1512	1 T30346	Chicken yes proto-
11	38.6	2.0	114955	1 X53491	Human adenosine A1
12	38.2	2.0	1593	1 T29199	Rat acetylglucosam
13	38.2	2.0	2247	1 Q37973	Human glycosyltran
14	37.4	2.0	429	1 Q69901	rAb Stage 1 VH con
15	37.4	2.0	2094	1 T33198	Alpha-1 A/D adrena
16	37	1.9	659	1 V15999	NBCCS (PTC) protei
17	37	1.9	3089	1 T05572	MEK33 cDNA. New si
18	37	1.9	3089	1 V22678	cDNA encoding a mi
19	37	1.9	3089	1 X56616	MEK33 protein codi
20	37	1.9	114955	1 X53491	Human adenosine A1
21	36.8	1.9	4258	1 Q22439	DNA of hEco-1, en
22	36.8	1.9	4258	1 T66529	Human C-type natri
23	36.8	1.9	6225	1 X55273	Human enzyme-relat
24	36.6	1.9	3841	1 T79967	Presenilin-interac
25	36.6	1.9	15872	1 T68715	Streptomyces venez
26	36.4	1.9	1417	1 V27893	Human hypothalamc
27	36.4	1.9	1929	1 Q43590	VCAM-6D/ICAM-1 DNA
28	36.4	1.9	1932	1 Q46661	VCAM-6D/ICAM-2 DNA
29	36.4	1.9	1941	1 Q43586	VCAM-6D DNA. Monoc
30	36.4	1.9	1941	1 Q46662	VCAM-6D/VCAM-1 DN
31	36.4	1.9	2091	1 Q23321	Soluble VCAM varia
32	36.4	1.9	2205	1 Q43587	VCAM/ICAM-1 DNA. M
33	36.4	1.9	2208	1 Q43588	VCAM/ICAM-2 DNA. M

34 36.4 1.9 2217 1 Q43585 VCAM-7D DNA. Monoc  
35 36.4 1.9 2220 1 Q10205 Sequence encoding  
36 36.4 1.9 2220 1 Q21005 VCAM variant with  
37 36.4 1.9 2220 1 Q29043 1E7-2G7 antigen cD  
38 36.4 1.9 2220 1 Q43589 VCAM/ICAM-3 DNA. M  
39 36.4 1.9 2220 1 Q43394 Sequence of bases  
40 36.4 1.9 2220 1 V16209 cDNA encoding huma  
41 36.4 1.9 2811 1 Q06687 vascular cell adhe  
42 36.4 1.9 3080 1 Q06688 vascular cell adhe  
43 36.2 1.9 3975 1 N81157 Malaria-specific g  
44 36.2 1.9 3975 1 Q22999 SERP gene. Recombi  
45 36.2 1.9 4020 1 T91361 Orf virus genomic

#### ALIGNMENTS

RESULT 1  
X28178  
ID X28178 standard; cDNA; 2040 BP.  
AC X28178;  
DT 16-JUN-1999 (first entry)  
DE Human PIGRL-1 coding sequence.  
KW PIGRL-1; human; autoimmune disease; hyper-IgM Immunodeficiency; HIM;  
KW X-linked Severe Combined Immunodeficiency; XSCID; IgA deficiency;  
KW diagnosis; therapy; ss.  
OS Homo sapiens.  
PN EP-905238-A2.  
PD 31-MAR-1999.  
PF 14-AUG-1998; 306487.  
PR 30-OCT-1997; US-961564.  
PR 25-AUG-1997; US-056935.  
PA (SMIK ) SMITHKLINE BEECHAM CORP.  
PI Sweet RW, Truneh A, Wu S;  
DR WPI: 99-192666/17.  
DT P-PSDB; Y05001.  
PT New polypeptides encoding PIGRL-1 useful for treating diseases such  
PT as X-linked Severe Combined Immunodeficiency  
PS Claim 4; Page 18-19; 26pp; English.  
CC This sequence encodes the human PIGRL-1 protein of the invention.  
CC Autoimmune diseases involving altered expression or activity of PIGRL-1  
CC may include Hyper-IgM Immunodeficiency (HIM), X-linked Severe Combined  
CC Immunodeficiency (XSCID) and IgA deficiency. These diseases can be  
CC diagnosed or susceptibility to them predicted by: (1) determining whether  
CC there is a mutation in the genomic copy of the gene encoding PIGRL-1; or  
CC (2) measuring the amount of PIGRL-1 in a sample derived from the patient.  
CC Patients deficient in PIGRL-1 can be treated by administering either the  
CC PIGRL-1 DNA or its complement or an agonist of PIGRL-1 to the patient.  
CC Patients with excessive expression or activity of PIGRL-1 can be treated  
CC by administering an antagonist of PIGRL-1, an antisenescence nucleic acid  
CC molecule which inhibits the expression of PIGRL-1 or administering  
CC sufficient PIGRL-1 to compete with the endogenous activity. PIGRL-1 can  
CC be used to identify its agonists by contacting a cell expressing PIGRL-1  
CC with a candidate compound in the presence of a signal system and noting  
CC the candidate as an agonist if a signal is produced. The same method can  
CC be used to identify antagonists of PIGRL-1 but the presence of an  
CC antagonist is indicated by a decrease in production of the signal.  
CC Antibodies against PIGRL-1 may be used to isolate or identify clones  
CC expressing PIGRL-1. Polynucleotides encoding PIGRL-1 may be used to  
CC identify chromosomal mutations in the gene encoding PIGRL-1 in patients.  
CC This information may then be correlated with the incidence of autoimmune  
CC disease in those patients to identify whether the mutation causes the  
CC disease.  
SQ Sequence 2040 BP; 478 A; 616 C; 494 G; 452 T;

Query Match 95.8%; Score 1830.4; DB 1; Length 2040;  
Best Local Similarity 99.1%; Pred. No. 0;  
Matches 1893; Conservative 0; Mismatches 11; Indels 6; Gaps 5;  
QY 5 aaaggaagcagcagcgtgcctccatcctctctagggcgtcttgatgacacctcactcta 64  
DB 77 GAGTAAGCAGCGTGTCTCCATCCCTCTCTAGGGGCTCTTGGAGGACCTTGCACACTCTA 136

QY 65 gaaggacaatggaacttctggcttggccacttacttcttgcagtatcagggccctg 124  
DB 137 GAAGGACAAATGGACTTCTGGCTTTGGCCACTTACTTCTGCCAGTATACGGGCCCTG 196  
QY 125 aggatctcccaagtaagtagaggggagctggcgagatcaggtaccatcaaatgc 184  
DB 197 AGGATCTCCACAGTAAGGTAGAGGGGAGCTGGGGGATCAGTACCATCAAGTGC 256  
QY 185 caacttccgaaatgaatgagagatatctgtgccggagatggctggatctgaaca 244  
DB 257 CCATCTCTGAATGCATGTGAGGATATCTGTGCCGGAGATGGCTGGATCGGAACA 316  
QY 245 tggtagctgagtgatctccaccacaaacttcaatcaaggcaagaatacaaggccaggttact 304  
DB 317 TGTGGTACCGTGGTATCCACCAACTTCATCAAGGCGAGATACAAAGGCGAGTACT 376  
QY 305 ctgaagcaataccacgcagaatctgttctagtggaggttaacacagctgcagaaagt 364  
DB 377 CTGAAGCAATACCCACGCAAGAATCTGTCTTAGTGGAGGTAACACAGCTGACAGAAAGT 436  
QY 365 gacagcgaagtctatgctgcgagcggagcaatacaacagaccgggaagaaccagaaa 424  
DB 437 GACAGCGAGTCTATGCTTGGGCGGATGAAACACAGACCGGGGAAAGACCCAGAAA 496  
QY 425 gtcacctgaatgtccacagtgaaacacagccatcatgggaagagcgcaatgctctgag 484  
DB 497 GTCACCTGAATGTCCACAGTAACGAGCCATCATGGGAAGAGACGCAATGCTGAG 556  
QY 485 actccaaatggttctatctgcttattgttccagatgctctgatatgcagttcttcc 544  
DB 557 ACTCCGAATGTTTCATCTGCTTATTTGTTCCAGATGCTGCTGATATGCTGCTTCC 616  
QY 545 aattgtcaacagagttaccacacagctcaagggggcaagggtccctcagttcaaccac 604  
DB 617 AAATTCGTAACACAGATTACACACAGCTCAAGGGGCAAGGTCCTCCAGTTCACAC 676  
QY 605 tcttcccccaaccccaaaacacccacgcctcagtggtccagagatcttcagtagca 664  
DB 677 TCCTCCCCACACCCCAATCAACCCACGCCCTCGAGTGTCCAGAGCATCTTCAGTAGCA 736  
QY 665 ggtgacaagcccaacacttctgcacatccatcacagctcaaaatctcagctcagag 724  
DB 737 GTGACACAGCCCGAGACCTTCTGCCATCCATACAGCTCAAAAATCTCAGCTCGAG 796  
QY 725 gggctgtcaagcccgagacgcccagctacaacacacacacacaggtcgcagggcagaga 784  
DB 797 GGGCTCTCAAGCCCCAGAGCCGCCACTACAAACACACACAGGCTGCACAGGCGAGAGA 856  
QY 785 gcactggactatggctcacagttgtgggaggaaggccaaggatttccacatcctgacctcg 844  
DB 857 GCACTGGACTATGGCTCACAGTCTGGAGGGAAGGCCAAGGATTTTCACATCCTGATCCG 916  
QY 845 accatctgggcttctctgtctggcaattctgggctggtggtgaaaggccgttgaa 904  
DB 917 ACCATCTGGGCTTTCTCTGTGGCACTTCTGGGGCTGGTGGTAAAGGGCGGTGAA 976  
QY 905 aggagaaagccctctccagcgccgcccgcgactggcgtgagatgcgcgccttgag 964  
DB 977 AGGAGGAAAGCCCTCTCCAGGGGGGCCGCCGACTGGCGCTGAGGATGCGCGCCCTGGAG 1036  
QY 965 agctcccaagagcccgaggttcgcgagaccgcgctcccaaaacacatctcacagccc 1024  
DB 1037 AGCTCCACAGAGCCCGCGGGTCCGCGGACCGCTCCCAAAACACATCTACACGCC 1096  
QY 1025 tgcccgcgcgcgtc-tggagcgcagctgcagcacagggggagggcccccgttcccgcc 1083  
DB 1097 TGCCCGCGCGCGCTCGTGGAGGGAGCGTGCAGGACACAGGGAGGAGGCCCGCTCCCGCG 1156  
QY 1084 cccgagagccgttgcccccccgccgctgcaggtgtcgtgatatccctggtccatgpc 1143  
DB 1157 CCGGAGAGCCGCTGGCCCCCGCCCGCTGCAGGTGCTTGAATCTCCCTGGCTCCATGCC 1216

QY 1144 ccattctgaagaccagctgtgaatacgtgagcctctaccaccagctgcgcgcatgatg 1203  
DB 1217 CCATCTCTGAAGACCAGCTGTGAATACGTGAGCCTCTACCACCAGCTGCGGCCATGATG 1276  
QY 1204 gaggacagtgattcagatgactacataatgttctgctgctgacaactccccagctatccc 1263  
DB 1277 GAGGACAGTGATTCAGATGACTACATCAATGTCTGCTGCACAACTCCCCAGCTATGCC 1336  
QY 1264 ccaaccccgagctcgagctgtggtgccaaggagttcatctatctgtgctgagtgccaatc 1323  
DB 1337 CCAACCCCAAGCTCGGACTGTGTGTGCCAAGGAGTCTCATCTATCTGCTGATGTCCAATAC 1396  
QY 1324 ctacttcattgtctcagagccctcacttcccatcccatcccatcctgcagctccatccc 1383  
DB 1397 CTGCTCATGTGTCTCAGAGCCCTCATCTCCCATGCCCATCTCGACTCCCATGCC 1456  
QY 1384 catctatctgtggccctgagcatggtctgtccccccaggtgcttcttgcacacctggcagc 1443  
DB 1457 CATCTATCTGT-GGCCCTGAGCATGGCTGTGCCCCAGGTGCTTGTGCACACTTGGCAGC 1515  
QY 1444 ccctgtatgtgacagtgaaagctgtagcattgtagagcaatgtcccaatggcaactgct 1503  
DB 1516 CCCCTGTAGTTGACAGGTAACTGTAGGCATGTAGAGCAATGTCCCAATGCCACTTGCT 1575  
QY 1504 tcccttccaaagccgtcgaacagactgtgggatttgcaagagtggttcttccatgtctttga 1563  
DB 1576 TCCTTTTCCAAAGCCGTCGAACAGACTGTGGGATTTTSCAGAGTGTTCCTTCCATGTCTTGA 1635  
QY 1564 ccaaggggtgtgtgtgtgcagagctctagatcacatggcatcagctggggcagagagca 1623  
DB 1636 CCACAGGGT-TGTTGCTGCCAGGGCTTAGATACATGATGATGATGATGATGATGATGATG 1694  
QY 1624 tagctatgtctcggggat -cettccaggggtggtgtcttcaacaataaagagctctg 1682  
DB 1695 TAGCTATGTCTCGGGCATCCCTTCCAGGGTGGGCTTTACACAAATGAGAGCTCTTG 1754  
QY 1683 ctctgagttatgtacgtgctcagcccccatggactaagcagggtctggtat--aaaca 1740  
DB 1755 CTCTGAGTTATGTGACATGCTCAGCCCCATGGACTAAGCAGGGGTCTGTGATATAAACA 1814  
QY 1741 ctctggaaacgcttggcctgatccaaatgttagcaactgtgtagtgaacgtctactta 1800  
DB 1815 CTCTGGAACGCTTGGCTTGCTGATCCAAATGTTAGCAGCTGTAGTGAAGCTTACTTA 1874  
QY 1801 tctcaagttctatgctaaaaggcaatttcttctgtgtagatgataaacaactattagc 1860  
DB 1875 TCTCAAGTCTATGTCTAAAGGCAATTTATCTGTGATGATGATGATGATGATGATGATGATG 1934  
QY 1861 aagatgcataatataccataaattcttcttactctgtctccatccttt 1910  
DB 1935 AAGATATGCATATATATCCATAAATTTCTTTTACTCTGTCTCCATCACTT 1984

RESULT 2  
X28179  
ID X28179 standard; cDNA; 1047 BP.  
AC X28179;  
DT 16-JUN-1999 (first entry)  
DE Human PIGRL-1 coding sequence fragment.  
KW PIGRL-1; human; autoimmune disease; hyper-IgM Immunodeficiency; HIM;  
KW X-linked Severe Combined Immunodeficiency; XSCID; IgA deficiency;  
OS diagnosis; therapy; ss.  
OS Homo sapiens.  
PN EP-905238-R2.  
PD 31-MAR-1999.  
PF 14-AUG-1998; 306487.  
PR 30-OCT-1997; US-961564.  
PR 25-AUG-1997; US-056935.  
PA (SMIK ) SMITHKLINE BEECHAM CORP.  
PI Sweet RW, Truneh A, Wu S;  
DR WPI: 99-192666/17.  
DR P-PSDB: Y05002.  
PT New polypeptides encoding PIGRL-1 useful for treating diseases such

PT as X-linked Severe Combined Immunodeficiency

FS Disclosure; Page 8; 26pp; English.  
 CC This sequence encodes the human PIGRL-1 protein of the invention.  
 CC Autoimmune diseases involving altered expression or activity of PIGRL-1  
 CC may include Hyper-IgM Immunodeficiency (HIM), X-linked Severe Combined  
 CC Immunodeficiency (XSCID) and IgA deficiency. These diseases can be  
 CC diagnosed or susceptibility to them predicted by: (1) determining whether  
 CC there is a mutation in the genomic copy of the gene encoding PIGRL-1; or  
 CC (2) measuring the amount of PIGRL-1 in a sample derived from the patient.  
 CC Patients deficient in PIGRL-1 can be treated by administering either the  
 CC PIGRL-1 DNA or its complement or an agonist of PIGRL-1 to the patient.  
 CC Patients with excessive expression or activity of PIGRL-1 can be treated  
 CC by administering an antagonist of PIGRL-1, an antisense nucleic acid  
 CC molecule which inhibits the expression of PIGRL-1 or administering  
 CC sufficient PIGRL-1 to compete with the endogenous activity. PIGRL-1 can  
 CC be used to identify its agonists by contacting a cell expressing PIGRL-1  
 CC with a candidate compound in the presence of a signal system and noting  
 CC the candidate as an agonist if a signal is produced. The same method can  
 CC be used to identify antagonists of PIGRL-1 but the presence of an  
 CC antagonist is indicated by a decrease in production of the signal.  
 CC Antibodies against PIGRL-1 may be used to isolate or identify clones  
 CC expressing PIGRL-1. Polynucleotides encoding PIGRL-1 may be used to  
 CC identify chromosomal mutations in the gene encoding PIGRL-1 in patients.  
 CC This information may then be correlated with the incidence of autoimmune  
 CC disease in those patients to identify whether the mutation causes the  
 CC disease.  
 SQ Sequence 1047 BP; 255 A; 281 C; 276 G; 223 T;

Query Match 35.4%; Score 676; DB 1; Length 1047;  
 Best Local Similarity 94.6%; Pred. No. 8e-176;  
 Matches 785; Conservative 0; Mismatches 34; Indels 11; Gaps 8;

QY 5 gagtaagcagcgtgtctccatcccccctctctagggcctttggatggacccttgactcta 64  
 DB 81 GAGTAAGCAGCGGTGTCTCCATCCCTCTCTAGGGGCTCTTGGATGGACCTTGCACCTTA 140  
 QY 65 gaaaggacaatggactcttgcttgccactttacttctctccagtatcaaggccctg 124  
 DB 141 GAAGGACAATGGACTTCTGGCTTGCCACTTTACTTCTGCCAGTATCAGGGGCCCTG 200  
 QY 125 aggatctccagaagtaaaagttagaggggaactggcgagatcattaccatcaaatgc 184  
 DB 201 AGGATCTCTCCAGAAGTAAAGGTAGAGGGGAGCTGGCGGATCAGTTACCATCAAGTGC 260  
 QY 185 ccacttctgaaatgcattgtgagatatactgtgccggagatgctggatctggaaca 244  
 DB 261 CCACTTCTGAAATGCAATGTGAGSATATATCTGTGCCGGAGATGGCTGGATCTGGAACA 320  
 QY 245 tctgtaccgtgtgtatccaccacacacttcatcaagcagaatacaaggccgagtaact 304  
 DB 321 TGTGTTACCTGTGTATCCACACCACTTCATCAAGCAGAGATACAAAGGCCCGAGTTACT 380  
 QY 305 ctgaagcaataccacacagaatctgttcttagtggaggttaacacagctgcagaaagt 364  
 DB 381 CTGAAGCAATACCACACAGAATCTGTCTTAGTGGAGGTAAACACAGCTGCAGAAAAGT 440  
 QY 365 gacagcagatctatgctcgcagcggcgatgaacacagacccgggaaagccacagaa 424  
 DB 441 GACAGCGAGTCTATGCTCTCGGA-CGGGCTGTAACACAGACCGGGGAAGACCCAGANA 499  
 QY 425 gtcaacctgaatgtccacagtgaatacagaccatcatgtggaagagagcagcgaatgctgag 484  
 DB 500 GTCACTCTGAAATGTCCACAGTGAATACAGCCATCATGGAAGAGAGAGAGCAATGCTGAG 559  
 QY 485 actcaaaatgttccatctgcctattgttccagatccctgcataatgcagttctcc 544  
 DB 560 ACTCAAAATGGTTTATCTGTGCTCTATTGTTCCAGATGCTGATATGCCGGTCTCTCC 619  
 QY 545 aaattcgtaac--cagagttaccacacacagc-tcaaggggcaaggtctccctcagttcac 601  
 DB 620 ACATTCTGTAACCGCAGATTACACACACAGCTTCAAGAGGGCAAGGTCTCTCCATTCTAC 679

QY 602 caactctctccccaccaccacaaa-tcaccaccgccc-tcagtgctccagagcatcttcag 659  
 DB 680 CACTCTCTCCCCACCACCACCAATTCACCCAGCCCTTCGAGTGTNCAGAGCATCTTCAG 739  
 QY 660 tagcaggtgacagcccg-aacctctctgcatccactacagcctcaaaaatctcagct 718  
 DB 740 TAGCAGGTGACAAGCCCGAAACTTTCTCTGCATCCACTACAGCTCAGAAATCTCAGCT 799  
 QY 719 ctggaggggctctcaagccccc--agacgcccagctacaaccacacacagcagctgcac 775  
 DB 800 CTGGAAGGCTCTTCAAGCCCGCAGAGCGCCAGCTACAA-CANCAACACAGGCTGCAC 858  
 QY 776 aggcagagagcctgagctactatgctcacagctctggaggggaagcccaag 825  
 DB 859 AGGCAGAGAGCACTGGATATTATGGGNTACAGCTCTGGGAGGGGAANG 908

RESULT 3  
 V62176  
 ID V62176 standard; DNA; 117213 BP.  
 AC V62176;  
 DT 13-JAN-1999 (first entry)  
 DE HSV-2 strain SB5; immunological response induction; therapy;  
 KW HSV-2 strain SB5; immunological response induction; therapy;  
 KW antiviral identification; viral protein inhibitor; ss.  
 OS Herpes simplex virus type 2.  
 FH Key location/Qualifiers  
 FT 755..1297  
 FT /tag= a  
 FT /product= "ORF#1 protein"  
 FT /note= "encoded protein shown in W72170"  
 FT 1170..2174  
 FT /tag= b  
 FT /product= "ORF#2 protein"  
 FT /note= "encoded protein shown in W72171"  
 FT 2229..2930  
 FT /tag= c  
 FT /product= "ORF#3 protein"  
 FT /note= "encoded protein shown in W72172"  
 FT complement (3130..3735)  
 FT /tag= d  
 FT /product= "ORF#4 protein"  
 FT /note= "encoded protein shown in W72173"  
 FT complement (3802..6447)  
 FT /tag= e  
 FT /product= "ORF#5 protein"  
 FT /note= "encoded protein shown in W72174"  
 FT 6017..8482  
 FT /tag= f  
 FT /product= "ORF#6 protein"  
 FT /note= "encoded protein shown in W72250"  
 FT 6026..8482  
 FT /tag= g  
 FT /product= "ORF#6f protein"  
 FT /note= "encoded protein shown in W72249"  
 FT 6065..8482  
 FT /tag= h  
 FT /product= "ORF#6e protein"  
 FT /note= "encoded protein shown in W72248"  
 FT 6167..8482  
 FT /tag= i  
 FT /product= "ORF#6d protein"  
 FT /note= "encoded protein shown in W72247"  
 FT 6296..8482  
 FT /tag= j  
 FT /product= "ORF#6c protein"  
 FT /note= "encoded protein shown in W72246"  
 FT 6326..8482  
 FT /tag= k  
 FT /product= "ORF#6b protein"  
 FT /note= "encoded protein shown in W72245"  
 FT 6446..8482  
 FT /tag= l

```
FT /product= "ORF#6a protein"
FT /transl_except= (pos: 7400. .7402, aa: Ala-Ala)
FT /transl_except= (pos: 7481. .7486, aa: Ile)
FT /note= "encoded protein shown in W72244"
FT 8457. .9347
CDS
FT /*tag= m
FT /product= "ORF#7 protein"
FT /note= "encoded protein shown in W72175"
FT complement (9604. .11855)
FT /*tag= n
CDS
FT /transl_except= (pos: 11635. .11636, aa: Ala)
FT /product= "ORF#8 protein"
FT /note= "encoded protein shown in W72176"
FT complement (11905. .14508)
FT /*tag= o
FT /product= "ORF#9b protein"
FT /note= "encoded protein shown in W72222"
FT complement (11905. .14520)
FT /*tag= p
CDS
FT /product= "ORF#9a protein"
FT /note= "encoded protein shown in W72223"
FT 14399. .15802
FT /*tag= q
FT /product= "ORF#10 protein"
FT /note= "encoded protein shown in W72177"
FT complement (15996. .16286)
FT /*tag= r
CDS
FT /product= "ORF#11 protein"
FT /note= "encoded protein shown in W72178"
FT complement (16202. .18064)
FT /*tag= s
FT /product= "ORF#12 protein"
FT /note= "encoded protein shown in W72179"
FT complement (18105. .19661)
FT /*tag= t
FT /product= "ORF#13 protein"
FT /note= "encoded protein shown in W72180"
FT complement (19415. .20074)
FT /*tag= u
CDS
FT /product= "ORF#14 protein"
FT /note= "encoded protein shown in W72181"
FT 20155. .21453
FT /*tag= v
FT /product= "ORF#15 protein"
FT /note= "encoded protein shown in W72182"
FT complement (21326. .22291)
FT /*tag= w
FT /product= "ORF#16 protein"
FT /note= "encoded protein shown in W72183"
FT complement (22546. .24654)
FT /*tag= x
FT /product= "ORF#17 protein"
FT /note= "encoded protein shown in W72184"
FT 24684. .25955
CDS
FT /*tag= y
FT /product= "ORF#18 protein"
FT /note= "encoded protein shown in W72185"
FT complement (26295. .27251)
FT /*tag= z
FT /product= "ORF#19 protein"
FT /note= "encoded protein shown in W72186"
FT complement (27630. .31754)
FT /*tag= aa
FT /product= "ORF#20b protein"
FT /note= "encoded protein shown in W72225"
FT complement (27630. .31784)
FT /*tag= ab
FT /product= "ORF#20a protein"
FT /note= "encoded protein shown in W72224"
FT complement (32067. .32735)
FT /*tag= ac
FT /product= "ORF#21 protein"
FT /note= "encoded protein shown in W72187"
FT 33140. .34984
FT /*tag= ad
FT /product= "ORF#22a protein"
FT /note= "encoded protein shown in W72226"
FT 33386. .34984
CDS
FT /*tag= ae
FT /product= "ORF#22b protein"
FT /note= "encoded protein shown in W72227"
FT complement (35205. .37721)
FT /*tag= af
FT /product= "ORF#23 protein"
FT /note= "encoded protein shown in W72188"
FT complement (38058. .39188)
FT /*tag= ag
FT /product= "ORF#24 protein"
FT /note= "encoded protein shown in W72189"
FT 39090. .39935
CDS
FT /*tag= ah
FT /product= "ORF#25 protein"
FT /note= "encoded protein shown in W72190"
FT 40216. .41973
FT /*tag= ai
FT /product= "ORF#26 protein"
FT /note= "encoded protein shown in W72191"
FT 42206. .44178
CDS
FT /*tag= aj
FT /transl_except= (pos: 44063. .44064, aa: Lys)
FT /product= "ORF#27 protein"
FT /note= "encoded protein shown in W72192"
FT complement (44853. .47297)
FT /*tag= ak
FT /product= "ORF#28 protein"
FT /note= "encoded protein shown in W72193"
FT 47122. .47338
FT /*tag= al
FT /product= "ORF#29 protein"
FT /note= "encoded protein shown in W72194"
FT complement (47305. .49662)
FT /*tag= am
FT /product= "ORF#30 protein"
FT /note= "encoded protein shown in W72195"
FT complement (50035. .51666)
FT /*tag= an
FT /product= "ORF#31 protein"
FT /note= "encoded protein shown in W72196"
FT complement (51701. .53575)
FT /*tag= ao
FT /product= "ORF#32 protein"
FT /note= "encoded protein shown in W72197"
FT 54393. .58115
CDS
FT /*tag= ap
FT /product= "ORF#33 protein"
FT /note= "encoded protein shown in W72198"
FT complement (58060. .58977)
FT /*tag= aq
FT /product= "ORF#34 protein"
FT /note= "encoded protein shown in W72199"
FT complement (58970. .60760)
FT /*tag= ar
FT /product= "ORF#35 protein"
FT /note= "encoded protein shown in W72200"
FT 60759. .61151
CDS
FT /*tag= as
FT /product= "ORF#36 protein"
FT /note= "encoded protein shown in W72201"
FT 61241. .62071
FT /*tag= at
FT /product= "ORF#37 protein"
FT /note= "encoded protein shown in W72202"
FT 62183. .62521
CDS
FT /*tag= au
FT /product= "ORF#38 protein"
FT /note= "encoded protein shown in W72203"
FT 62183. .62521
```

















```

Query Match      2.0%; Score 37.4; DB 1; Length 429;
Best Local Similarity 53.8%; Pred. No. 0.64; 56; Indels 0; Gaps 0;
Matches 77; Conservative 0; Mismatches 0;

QY 285 aatacaaggccgagttactctgaagcaataccccacgcagaatctgttcttagtgaggg 344
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 245 AGTTCAGGTCAGAGTGACAATGCTGTGTAGACACGACGACGACGACGACGACGACGAC 304
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 345 taacacagctgacagaagtagacacgagagctctatgctcgagcgcgagcatgaacacag 404
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 305 TCAGCAGCCTGACGCCGCCGACGCCGCCGCTATTATTGTGCAGACGGAATGCGGTAT 364
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 405 accggggaagaccagaaagtc 427
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 365 CAACGGGATATGCTCTGGACTTC 387
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
T33198/c
ID T33198 standard; cDNA to mRNA; 2094 BP.
AC T33198;
DT 06-NOV-1996 (first entry)
DE Alpha-1 A/D adrenaline receptor coding sequence.
KW Alpha-1 A/D adrenaline receptor; agonist; antagonist; detection;
KW screening; recombinant production; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Key 22..1740
FT cds /*tag= a
FT /product= alpha-1 A/D adrenaline receptor
FT
PN J07289265-A.
PD 07-NOV-1995.
PF 28-APR-1994; 091041.
PR 28-APR-1994; JP-091041.
PA (NNSH ) NIPPON SHINYAKU CO LTD.
DR WPI; 96-015270/02.
DR P-PDB; R98563.
PT DNA encoding alpha-1A/D adrenaline receptor - useful for
PT identification of (ant)agonist(s)
PS Disclosure; ; lpp; Japanese.
CC The present sequence encodes a human alpha-1 A/D adrenaline receptor.
CC The DNA can be used to express the receptor in large quantities by
CC recombinant DNA technology. (Ant)agonists selective to the receptor
CC can also be detected.
SQ Sequence 2094 BP; 305 A; 726 C; 707 G; 356 T;

Query Match      2.0%; Score 37.4; DB 1; Length 2094;
Best Local Similarity 48.4%; Pred. No. 1.4;
Matches 104; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 915 cctctccagggcgccgcgagctggcgtgagatggcgccctggagagctccacaga 974
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 238 CCGGCTCCCCCGGAGAGTCCGGTGTCTCTGCCGCTGCTGCCACACGCGCGCGC 179
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 975 ggcctcgagggtgcgcgcgagcgcctcccaaaacacatctacagcgcttcgcgcgcgc 1034
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 178 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 119
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1035 gcgctctgagggagcgagcgagcgagggagggagggcgcttcgcgcgcgcgcgcgcgc 1094
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 118 CCGCGCTGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 59
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1095 gttgcccccccgcccgctgcaggtgtctgaatctc 1129
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 58 CCTCGAAACTGACGCTCAGGAGATCGCGGAAAGTC 24
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: September 12, 2000, 11:58:00  
 Job time: 6521 sec



OM of: US-09-135-238B-2\_COPY\_273\_390 to: EST:\* out\_format : pfs  
 Date: Sep 12, 2000 6:04 AM  
 About: Results were produced by the GenCore software, version 4.5,  
 Copyright (c) 1993-2000 CompuGen Ltd.

## Command line parameters:

-MODEL=frame+pn2n.model -DEV=xlp  
 -O=/sgn2\_1/USPFO\_spool/US09135238/runat\_24082000\_091323\_17869/app\_query.fasta\_1.1282  
 -B=EST -QFWI=fastcap -SUFFI=phn.rst -GAPOP=12.000 -GAPEXT=4.000  
 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -GAPOP=4.500  
 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human4.0.cdi  
 -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0  
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0  
 -MAXLEN=1000000 -USER=US09135238 -CGN1\_1\_2178 -NCPU=6 -ICPU=3  
 -LONGLOG -NO\_XLUPXY -WAIT -THREADS=1

## Search information block:

Query: US-09-135-238B-2\_COPY\_273\_390  
 Query length: 118  
 Database: EST:\*  
 Database sequences: 5247842  
 Database length: -2090053206  
 Search time (sec): 1563.530000

## score\_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
gb_est25:AI760373	-	610.00	823.24	711	AI760373 w87d11.x1 NCI_CGAP_CL1
gb_est7:AA456707	+	551.00	748.58	7.0e-33	AA456707 aal3g04.r1 Soares_NHMH
gb_est42:AA401870	+	372.00	507.30	2.0e-19	AI401870 UI-HF-BK0-aah-g-11-0-0
gb_est21:AI508502	+	331.50	449.96	3.1e-16	AI508502 vb34f04.y1 Soares_mous
gb_est4:AA290194	+	251.00	343.11	2.8e-10	AA290194 vb34f04.y1 Soares_mous
gb_est21:AI468023	+	210.00	288.26	3.2e-07	AI468023 t384h03.x1 Soares_NSF
gb_est24:AI743020	+	210.00	287.01	3.7e-07	AI743020 w85b10.x1 Soares_NSF
gb_est16:AI137500	-	178.00	246.20	7.0e-05	AI137500 UI-R-C2p-oc-a-09-0-UI
gb_est3:AA174968	+	142.00	196.46	0.0410	AA174968 ms78h03.r1 Soares_mous
gb_est30:AV057274	+	111.00	159.02	4.99	AV057274 AV057274 Mus musculus
gb_est46:AW727502	+	106.50	140.61	52.91	AW727502 GA_Ea0012x08 Gossyp
gb_est46:AW727557	+	102.50	133.52	131.29	AW727557 GA_Ea0012022 Gossyp
gb_est46:AW731188	+	102.50	133.17	137.39	AW731188 GA_Ea0010115 Gossyp
gb_est46:AW731177	+	102.00	134.33	118.34	AW731177 GA_Ea0010617 Gossyp
gb_est46:AW731156	+	100.50	130.88	184.08	AW731156 GA_Ea0010C23 Gossyp
gb_est18:AI288070	+	100.00	140.99	50.39	AI288070 qv70c08.x1 NCI_CGAP_U
gb_est39:AW187330	+	100.00	138.14	72.57	AW187330 BNLG112883 Six-day C
gb_est36:AW1091361	+	99.50	139.88	58.07	AI091361 oc02f11.x1 Soares_NSF
gb_est38:AW170453	+	99.50	138.39	70.27	AW170453 xn61h01.x1 Soares_NHMH
gb_est46:AW727550	+	99.00	130.66	189.41	AW727550 GA_Ea0012008 Gossyp
gb_est15:AI045027	+	98.50	138.49	69.38	AI045027 UI-R-C1-kf-c-01-0-UI
gb_est26:AI857947	+	98.50	134.09	122.08	AI857947 w168e01.x1 NCI_CGAP_U
gb_gss1:AAQ12086	-	98.50	128.57	247.77	AAQ12086 520A063071497 Cosm
gb_est31:AV140597	-	98.00	141.24	48.80	AV140597 AV140597 Mus musculus
gb_est42:AW643326	+	98.00	136.80	86.15	AW643326 BP230015B10A6 Soares
gb_est38:AW123443	+	97.50	137.41	79.72	AW123443 UI-M-BH2.1-apj-d-12-0
gb_est36:AV394694	-	97.50	135.90	96.71	AV394694 AV394694 Chlamydomon
gb_est40:AW287186	-	97.00	134.96	109.17	AW287186 LGL_267_E12_b1_A002_L
gb_est46:AW731186	+	96.50	127.41	287.34	AW731186 GA_Ea0010111 Gossyp
gb_est46:AW727510	+	96.50	126.69	315.23	AW727510 GA_Ea0012K24 Gossyp
gb_est46:AW044901	+	96.00	132.14	156.69	AW044901 um15b10.y1 Sugano_mous
gb_gss12:AAQ833643	+	96.00	131.68	166.33	AAQ833643 nbx0024J12r CUGI Rice
gb_est46:AW731195	+	96.00	122.99	506.42	AW731195 GA_Ea0010K05 Gossyp
gb_est24:AI742962	+	95.50	136.27	92.28	AI742962 w876d09.x1 Soares_NSF
gb_est5:AA309971	+	95.50	134.78	111.74	AA309971 EST180822 Jurkat T-cell
gb_est20:AI419942	-	95.50	133.82	126.38	AI419942 t340d11.x1 Soares_NFL
gb_est17:AI220032	-	95.50	132.40	151.51	AI220032 qg78h02.x1 Soares_NFL
gb_est46:AW727500	+	95.50	125.56	364.48	AW727500 GA_Ea0012K04 Gossyp
gb_est46:AW731153	+	95.50	123.33	375.35	AW731153 GA_Ea0010C17 Gossyp
gb_est26:AI845713	+	95.00	134.48	116.00	AI845713 UI-M-AQ1-ady-h-03-0-UI
gb_est26:AI839944	-	95.00	133.88	125.39	AI839944 UI-M-AH0-acw-h-12-0-UI

gb\_est39:AW239576 + 95.00 131.62 167.51 550 ! AW239576 pt1lc.pk001.k19 chi  
 gb\_est46:AW727477 + 95.00 123.48 475.77 1429 ! AW727477 GA\_Ea0012I06 Goss  
 gb\_est46:AW727558 - 95.00 122.02 574.14 1697 ! AW727558 GA\_Ea0012024 Goss  
 gb\_gss12:AAQ858381 - 94.50 131.79 163.84 498 ! AQ858381 nbeb0012H11f CUGI R

seq\_name: gb\_est25:AI760373

## seq\_documentation\_block:

LOCUS AI760373 711 bp mRNA EST 20-DEC-1999  
 DEFINITION wh87d11.x1 NCI\_CGAP\_CL1 Homo sapiens cDNA clone IMAGE:2387733 3'  
 similar to TR:060667 060667 ANTI-FAS-INDUCED APOPTOSIS. ; mRNA  
 sequence.  
 ACCESSION AI760373  
 VERSION AI760373.1 GI:5176040  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 711)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On Oct 30, 1998 this sequence version replaced gi:3812483.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert\_Strausberg@nih.gov  
 Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever,  
 M.D., Louis M. Staudt, M.D., Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D.  
 DNA Sequencing by: Greg Lennon, Ph.D.  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 www-bio.llnl.gov/borip/image/image.html  
 Insert Length: 1449 Std Error: 0.00  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 472.

## FEATURES

Location/Qualifiers  
 1..711  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2387733"  
 /clone.lib="NCI\_CGAP\_CL1"  
 /tissue.type="B-cell, chronic lymphocytic leukemia"  
 /lab.host="DH10B"  
 /note="vector: p7T3p-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
 was primed with a Not I - oligo(dT) primer [5'  
 TGTACCAATCTGAAGTGGAGCGCGCATGCTTTTCTTTTCTTTTCTTTT  
 T 3']; double-stranded cDNA was ligated to Eco RI  
 adaptors (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of the modified p7T3 vector.  
 Library is normalized, and was constructed by Bento  
 Soares and M. Fatima Bonaldo."

BASE COUNT 127 a 195 c 239 g 148 t 2 others  
 ORIGIN

## alignment\_scores:

Quality: 610.00 Length: 118  
 Ratio: 5.214 Gaps: 0  
 Percent Similarity: 99.153 Percent Identity: 99.153

## alignment\_block:

US-09-135-238B-2\_COPY\_273\_390 x AI760373/rev ...

Align seg 1/1 to reverse of: AI760373 from: 1 to: 711

1 LysArgAlaValGluArgArgLysAlaLeuSerArgAlaArgLeu 17

461 AAAAGGCGCGTGAAGAGGAGGAAGCCCTCCAGCGCGCGCGGACT 412

```

17 uAlaValArgMetArgAlaLeuGluSerSerGlnArgProArgGlySerP 34
|||||
411 GCGCTGAGGATGCGCGCCCTGGAGAGCTCCAGAGGCCCGCGGGTCGC 362
34 roArgProArgSerGlnAsnAsnLeuYrSerAlaCysProArgArgAla 50
|||||
361 CGCGAGCGCGCTCCCAAAACAATCTACAGCGCTTGCOCGCGCGCGCT 312
51 ArgGlyAlaAspAlaAlaGlyThrGlyGluAlaProValProGlyProG1 67
|||||
311 CTGTGAGCGGAGCTGCAGGCACAGGAGGCCCGCTTCCGGGCCCGG 262
67 yAlaProLeuProProAlaProAlaProLeuGlnValSerGluSerProTrpLeuH 84
|||||
261 AGCGCGCTTGCOCGCGCGCGCTGCAGGTTGTTGAATCTCCTGGCTCC 212
84 isAlaProSerLeuYrThrSerCysGluYrValSerLeuYrHisGln 100
|||||
211 ATGCCCATCTCTGAGACCACTGTGAATACGTGAGCCTTACCACAG 162
101 ProAlaAlaMetMetGluAspSerAspSerAspAspTyrIleAsnValPr 117
|||||
161 CTGCGCGCATGATGGAGGACAGTGAATCAGATGACTACATCAATGTTC 112
117 oAla 118
||||
111 TGCC 108

seq_name: gb_est7:AA456707

seq_documentation_block:
LOCUS AA456707 404 bp mRNA EST 06-JUN-1997
DEFINITION 2a13904.r1 Soares_NHMPu_S1 Homo sapiens CDNA clone IMAGE:813174
5', mRNA sequence.
ACCESSION AA456707
VERSION AA456707.1 GI:2179283
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 404)
Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,
Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
White,Y., Wylie,T., Waterston,R. and Wilson,R.
WashU-Merck EST Project 1997
Unpublished (1997)
On Jan 14, 1998 this sequence version replaced gi:1877745.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL: contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28m3 rev2 EF from Amersham
High quality sequence stop: 390.
Location/Qualifiers
1. .404
/organism="Homo sapiens"
/db_xref="GDB:6044084"
/db_xref="taxon:9606"
/clone_lib="IMAGE:813174"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/notes="Organ: mixed (see below); Vector: pT7B3D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; Equal amounts of plasmid DNA from three

```

normalized libraries (melanocyte 2NBHM, pregnant uterus NBHPU, and fetal heart NBH19W) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 260232-265223, 340488-345479, and 484488-489479."

BASE COUNT 71 a 153 c 114 g 66 t

ORIGIN

alignment\_scores:  
Quality: 551.00 Length: 111  
Ratio: 5.009 Gaps: 0  
Percent Similarity: 99.099 Percent Identity: 98.198

alignment\_block:  
US-09-135-238B-2\_COPY\_273\_390 x AA456707 ..

Align seg 1/1 to: AA456707 from: 1 to: 404

```

8 LysAlaLeuSerArgArgAlaArgArgLeuAlaValArgMetArgAlaLe 24
:::|||||
1 GAAGCCCTCTCCAGCGCGCGCGGCTGCGGAGTGCAGGATGCGCGCCT 50
24 uGluSerSerGlnArgProArgGlySerProArgProArgSerGlnAsnA 41
|||||
51 GGAGAGCTCCAGAGGCCCGCGGGTCCGCGGA.CCGCGCTCCCAAAACA 99
41 snIleYrSerAlaCysProArgArgAlaArgGlyAlaAspAlaAlaGly 57
|||||
100 ACATCTACAGCGCTGCCGCGCGGCTCTGTGGAGCGGACGCTGCAGGC 149
58 ThrGlyGluAlaProValProGlyProGlyAlaProLeuProProAlaPr 74
|||||
150 ACAGGGGAGGCCCGCGGTTCGCGCGCGGAGCGCGGTGCCCGCGCC 199
74 oLeuGlnValSerGluSerProTrpLeuHisAlaProSerLeuYrThrS 91
|||||
200 GCTCAGGTGTGTGAATCTCCCTGGCTCCATGCCCATCTCTGAAGACCA 249
91 erCysGluYrValSerLeuYrHisGlnProAlaAlaMetMetGluAsp 107
|||||
250 GCTGTGAATACGTGAGCTCTACACAGCGCTGCCGCCATCATGAGGAG 299
108 SerAspSerAspAspTyrIleAsnValProAla 118
|||||
300 AGTGATTTCAGATGACTACATCAATGTTCTCTGCC 332

```

seq\_name: gb\_est42:AA401870

seq\_documentation\_block:  
LOCUS AW401870 410 bp mRNA EST 16-FEB-2000  
DEFINITION UI-HF-BK0-aah-g-11-0-UI.r1 NIH\_MGC\_36 Homo sapiens CDNA clone  
IMAGE:3053781 5', mRNA sequence.

ACCESSION AW401870  
VERSION AW401870.1 GI:6920556  
KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 410)

REFERENCE NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE Unpublished (1999)

JOURNAL On Jun 22, 1998 this sequence version replaced gi:3246884.

COMMENT

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Eco RI site shown at the beginning of the sequence.

Tissue Procurement: Louis M. Staudt, M.D., Ph.D.

CDNA Library Preparation: M.B. Soares Lab  
 CDNA Library Arrayed by: M.B. Soares Lab  
 DNA Sequencing by: M.B. Soares Lab  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 www-bio.llnl.gov/bbrp/image/image.html  
 Seq primer: M13 Forward

# FEATURES

Location/Qualifiers  
 1..410

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3053781"  
 /clone\_lib="NIH\_MGC\_36"  
 /tissue\_type="lymph"  
 /cell\_type="germinal center B cells"  
 /cell\_line="MGC85"  
 /lab\_host="DH10B (LTI)"  
 /note="Vector: pT73-Pac; Site\_1: NotI; Site\_2: Eco RI;  
 Constructed from size fractionated cytoplasmic mRNA  
 (0.5-1.5kb). Directionally cloned. Cells provided by Louis  
 M. Staudt, Ph.D. Library preparation by Maria de Fatima  
 Ronaldo, Ph.D. and M. Bento Soares, Ph.D."

BASE COUNT 72 a 154 c 94 g 89 t 1 others

## ORIGIN

alignment\_scores:  
 Quality: 372.00 Length: 69  
 Ratio: 5.391 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-135-238B-2\_COPY\_273\_390 x AW401870

Align seg 1/1 to: AW401870 from: 1 to: 410

50 AlaArgGlyAlaAspAlaGlyThrGlyGluAlaProValProGlyPr 66  
 10 GCTCGTGGAGCGAGCGCTGCAGGCACAGGGAGGCGCCGCTCCGGGCC 59  
 66 oGlyAlaProLeuProAlaProLeuGlnValSerGluSerProTrpL 83  
 60 CGAGCGCGCTTGGCCCGCCGCGCTGCAGGTGTCTGAATCTCCGCGC 109  
 83 euHisAlaProSerLeuLysThrSerCysGluTyrValSerLeuTyrHis 99  
 110 TCCATGCCCATCTCTGAAGACAGCTGTGAATACGTGAGCGCTTACCAC 159  
 100 GlnProAlaAlaMetMetGluAspSerAspSerAspTyrIleAsnVa 116  
 160 CAGCGTCCGCCATCATGGAGGACAGTATTCAGATGACTACATCATGT 209

116 lProAla 118  
 210 TCCTGCC 216

seq\_name: gb\_est21:AI508502

seq\_documentation\_block:  
 LOCUS AI508502 568 bp mRNA EST 12-MAR-1999  
 DEFINITION VB34f04.y1 Soares mouse lymph node NBMLN Mus musculus cDNA clone  
 IMAGE:750847 5', mRNA sequence.

ACCESSION AI508502

VERSION AI508502.1 GI:4407407

## KEYWORDS

house mouse.

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 568)  
 Marra, M., Hallier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,  
 Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,  
 Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,

Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,  
 Waterston, R. and Wilson, R.  
 The WashU-NCI Mouse EST Project 1999  
 Unpublished (1999)  
 On Nov 2, 1998 this sequence version replaced gi:3830418.  
 Contact: Marra M/WashU-NCI Mouse EST Project 1999  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.edu

# TITLE

JOURNAL

## COMMENT

This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:459831

This read is a RESSEQUENCE of a previously sequenced mouse clone  
 This read has been verified (found to hit its original self in the  
 correct orientation)

Putative full length read  
 Vector to vector length is 639  
 Seq primer: -40RP from Gibco  
 High quality sequence stop: 458.

## FEATURES

source

1..568  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:750847"  
 /sex="male"  
 /tissue\_type="lymph node"  
 /dev\_stage="4 weeks"  
 /lab\_host="DH10B"  
 /note="Organ: lymph node; Vector: pT73D-Pac (Pharmacia)  
 with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
 1st strand cDNA was primed with a Not I - oligo(dT) primer  
 [5'  
 TGTACCAATCTGAAGTGGAGCGCGGATACATTTTTTTTTTTTTTTTTTTTT  
 3']; double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pT73 vector. RNA  
 provided by Dr. Bertrand Jordan. Library constructed and  
 normalized by Bento Soares and M. Fatima Ronaldo."

BASE COUNT 109 a 191 c 144 g 124 t

## ORIGIN

alignment\_scores:  
 Quality: 331.50 Length: 115  
 Ratio: 3.900 Gaps: 1  
 Percent Similarity: 73.913 Percent Identity: 60.000

## alignment\_block:

US-09-135-238B-2\_COPY\_273\_390 x AI508502

Align seg 1/1 to: AI508502 from: 1 to: 568

8 LysAlaLeuSerArgArgAlaArgArgLeuAlaValArgMetArgAlaLe 24  
 42 GAAGCCCTCTCCAGAGCTGGGGCCGACTAGCGATGAGGAGCGGAGCGG 91  
 24 uGluSerSer.....GlnArgProArgGlySerProArgp 36  
 92 GGGGGCTTCGCGCGCTTCCCCACACAGCGCGGGATGCTCGCAGAGGC 141  
 36 roArgSerGlnAsnAsnIleTyrSerAlaCysProArgArgAlaArgGly 52  
 142 CGCGCTCGAGACACAGCTACAGCGCTGCCCGCGCGCGCAGCGGGA 191  
 53 AlaAspAlaGlyThrGlyGluAlaProValProGlyProGlyAlaPr 69  
 192 CCAGACAGCTGGGTCCAGCGGAGGTCCGCTCCTCAGCGCCCGCCAGCTC 241  
 69 oLeuProAlaProLeuGlnValSerGluSerProTTrpLeuHisAlaP 86





Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 5 libraries. The pools consisted of the following libraries and clones: Soares NBHSF pool 1: 309384-310919, 323208-325895 Soares NB2HP pool 1: 145032-147335, 147720-148103, 148872-149255, 15002 - 150407, 151176-152327 Soares NB2HF8-9W pool 1: 758280-760583, 772104-774407 Soares NBHPA pool 1: 304776-306311, 320136-322823, 326280-326663 Soares NBHOT pool 1: 723720-726407, 739080-740999 Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 123 a 97 c 99 g 131 t  
ORIGIN

alignment\_scores:  
Quality: 210.00 Length: 38  
Ratio: 5.526 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-135-238B-2\_COPY\_273\_390 x AI468023 ..  
Align seg 1/1 to: AI468023 from: 1 to: 450

81 ProTrpLeuHisAlaProSerLeuLysThrSerCysGluTyrValSerLe 97  
|||||  
321 CCTCGGCTCCATGCCCATCTCTGAAGACGAGCTGTGAATACGTGAGCCT 370  
97 uTyrHisGlnProAlaAlaMetMetGluAspSerAspSerAspTyrI 114  
|||||  
371 CTACCACGCTCCGCCCATGATGGAGGACAGTGCATTGAGTACTACA 420  
114 leAsnValProAla 118  
|||||  
421 TCAATGTTCTGTCGCC 434  
seq\_name: gb\_est24:AI743020

seq\_documentation\_block:  
LOCUS AI743020 521 bp mRNA EST 21-JUN-1999  
DEFINITION wg85b10.x1 Soares NSF\_F8\_9W\_OT\_PA\_P\_S1 Homo sapiens cDNA clone IMAGE:2371867 3' similar to TR:060667 060667 ANTI-FAS-INDUCED APOPTOSIS. ; mRNA sequence.  
ACCESSION AI743020  
VERSION AI743020.1 GI:51111308  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 521)  
AUTHORS NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LLNL; contact the IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Seq primer: -400P from Gibco  
High quality sequence stop: 462.  
Location/Qualifiers  
1..521  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2371867"  
/clone\_lib="Soares\_NSF\_F8\_9W\_OT\_PA\_P\_S1"  
/lab\_host="DH10B"  
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;

FEATURES  
source

Equal amounts of plasmid DNA from five normalized libraries were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 5 libraries. The pools consisted of the following libraries and clones: Soares NBHSF pool 1: 309384-310919, 323208-325895 Soares NB2HP pool 1: 145032-147335, 147720-148103, 148872-149255, 15002 - 150407, 151176-152327 Soares NB2HF8-9W pool 1: 758280-760583, 772104-774407 Soares NBHPA pool 1: 304776-306311, 320136-322823, 326280-326663 Soares NBHOT pool 1: 723720-726407, 739080-740999 Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 137 a 122 c 115 g 147 t  
ORIGIN

alignment\_scores:  
Quality: 210.00 Length: 38  
Ratio: 5.526 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-135-238B-2\_COPY\_273\_390 x AI743020 ..  
Align seg 1/1 to: AI743020 from: 1 to: 521

81 ProTrpLeuHisAlaProSerLeuLysThrSerCysGluTyrValSerLe 97  
|||||  
319 CCTCGGCTCCATGCCCATCTCTGAAGACGAGCTGTGAATACGTGAGCCT 368  
97 uTyrHisGlnProAlaAlaMetMetGluAspSerAspSerAspTyrI 114  
|||||  
369 CTACCACGCTCCGCCCATGATGGAGGACAGTGCATTGAGTACTACA 418  
114 leAsnValProAla 118  
|||||  
419 TCAATGTTCTGTCGCC 432  
seq\_name: gb\_est16:AI137500

seq\_documentation\_block:  
LOCUS AI137500 398 bp mRNA EST 05-JUL-1999  
DEFINITION UI-R-C2p-oc-a-09-0-UI-s1 UI-R-C2p Rattus norvegicus cDNA clone UI-R-C2p-oc-a-09-0-UI 3', mRNA sequence.  
ACCESSION AI137500  
VERSION AI137500.1 GI:3638277  
KEYWORDS EST.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
REFERENCE 1 (bases 1 to 398)  
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene discovery  
JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 97044477  
COMMENT On Apr 3, 1998 this sequence version replaced gi:3018677.  
Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
University of Iowa  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: msoares@blue.wesg.uiowa.edu

The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized adult spleen library. cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics This clone is also available through the I.M.A.G.E.



54 CCTGATTTCAGATGATTACATCAATATTTCCT 84

[illegible]





```

    Quality: 100.50      Length: 95
    Ratio: 2.138         Gaps: 3
    Percent Similarity: 49.474      Percent Identity: 36.842

alignment_block:
US-09-135-238B-2_COPY_273_390 x AW7311156  ..

Align seq 1/1 to: AW7311156 from: 1 to: 1430

2  ArgAlaValGluArgArgTysAlaLeuSerArgArgAlaArgArgLeuAl 18
897  CGCTCGGCGCGCCGCCGCCGCCGCGCTCCGCCGCGCGCGCGCGCGCC 946
18 aVal.ArgMetArgAlaLeuGlu...SerSerGlnArgProArgTys 33
947  GCGCGCGCGCGCGCGCTCGGCGCCGCCGCCGCGCGCTCCGCCGCGCGC 996
34  Pro..... 34

997  GCGACACCGCGCGCGCGCGCGCGCGCTCCGCCGCCGCCGCCGCGCTGGG 1046
35  .....ArgProArgSerGlnAsnAsnIleTyrSerAlaCysp 47
      |||||.....:
1047  CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCCC 1096
47  roArgArgAlaArgGlyAlaAspAlaAlaGlyThrGlyGluAlaProVal 63
      ||||| ||| ||||| ||||| ||||| ||||| ||||| |||||
1097  CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 1146
64  ProGlyProGlyAlaProLeuProProAlaPro 74
      ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1147  CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 1179

```

OM of: US-09-135-238B-2\_COPY\_18\_272 to: N\_Geneseq\_36.\* out\_format : pfs

Date: Sep 12, 2000 6:56 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-O=/cnp2.1/USPTO\_spool/US09135238/runat\_24082000\_091324\_17918/app\_query.fasta\_1.1282  
-DB=N\_Geneseq\_36 -QFMT=fastap -SUFFIX=p2n.rng -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -GAPOP=10.000 -XGAPEXT=0.500  
-GAPOP=6.000 -GAPEXT=7.000 -GAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosome2  
-TRANS=human40.cdi -LIST=45 -DLOCALIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTPMT=pfs  
-NORM=ext -MINLEN=0 -MAXLEN=1000000  
-USER=US09135238 -CGN1\_1\_123 -NCPU=6 -ICPU=3 -LONGLOG -NO\_XLPXY  
-WAIT -THREADS=1

Search information block:

Query: US-09-135-238B-2\_COPY\_18\_272  
Query length: 255  
Database: N\_Geneseq\_36.\*  
Database sequences: 311585  
Database length: 125096042  
Search time (sec): 121.040000

score\_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
N_Geneseq_36:X28178	+ 1340.00	2707.38	4.1e-108	2040	Human PIGRL-1 coding sequence.
N_Geneseq_36:X28179	+ 1034.00	1605.11	8.4e-82	1047	Human PIGRL-1 coding sequence.
N_Geneseq_36:T31290	+ 197.00	284.24	3.1e-08	3095	Mouse poly-Ig receptor protein
N_Geneseq_36:T31291	+ 188.00	247.67	3.4e-06	29392	Mouse poly-Ig receptor protein
N_Geneseq_36:T31292	+ 185.50	265.70	3.4e-07	3269	Rat poly-immunoglobulin receptor
N_Geneseq_36:T31293	+ 180.50	263.62	4.4e-07	1839	Human secretory immunoglobulin
N_Geneseq_36:T31294	+ 180.50	263.11	4.7e-07	1936	Secretory component DNA fragment
N_Geneseq_36:T31295	+ 180.50	262.63	5.0e-07	2031	Secretory component DNA fragment
N_Geneseq_36:T31296	+ 180.50	259.00	8.0e-07	2919	Human poly-immunoglobulin receptor
N_Geneseq_36:T31297	+ 175.50	255.26	1.3e-06	3630	Bovine poly-immunoglobulin receptor
N_Geneseq_36:T31298	+ 175.50	255.41	1.3e-06	1911	CDNA for human immunity related
N_Geneseq_36:T31299	+ 166.50	241.51	7.5e-06	1876	Clone 1 for truncated poly Ig
N_Geneseq_36:T31300	+ 166.50	235.23	1.7e-05	3517	Rabbit poly-immunoglobulin receptor
N_Geneseq_36:T31301	+ 145.00	207.86	0.0006	1876	Clone 2 for truncated poly Ig
N_Geneseq_36:T31302	+ 122.00	176.86	0.0299	1137	Human PIGR-2 coding sequence.
N_Geneseq_36:T31303	+ 118.50	165.33	0.1313	2084	Rat kidney injury related molecule
N_Geneseq_36:T31304	+ 118.50	163.25	0.1714	2566	Rat kidney injury related molecule
N_Geneseq_36:T31305	+ 117.50	165.26	0.1325	1795	Human kidney injury related molecule
N_Geneseq_36:T31306	+ 116.00	174.46	0.0407	565	EST clone EN260. New polynucleotide
N_Geneseq_36:T31307	+ 107.50	154.93	0.4982	1053	Human secreted protein gene
N_Geneseq_36:T31308	+ 107.50	154.73	0.5115	1075	Human secreted protein gene
N_Geneseq_36:T31309	+ 106.50	152.89	0.6477	1105	Human secreted protein gene
N_Geneseq_36:T31310	+ 104.50	129.15	13.59	8690	Plasmid pD17-hm14.H1. Inhibitor
N_Geneseq_36:T31311	+ 104.00	146.42	1.48	1427	Plasmid Hu19BHCpcd encoding an
N_Geneseq_36:T31312	+ 103.00	144.85	1.81	1427	Plasmid Hu19BHCpcd encoding an
N_Geneseq_36:T31313	+ 103.00	144.85	1.81	1427	Plasmid Hu19BHCpcd encoding an
N_Geneseq_36:T31314	+ 102.50	126.46	19.21	8321	Plasmid expressing hBR96-A. In
N_Geneseq_36:T31315	+ 102.00	128.05	15.65	6557	Anti-IgE VH expression vector.
N_Geneseq_36:T31316	+ 101.50	134.47	6.87	3390	CD22-beta gene. Inhibition of
N_Geneseq_36:T31317	+ 101.00	141.69	2.72	1431	Primate anti-human B7.1 ant
N_Geneseq_36:T31318	+ 101.00	141.69	2.72	1431	Macaque primatized 7C10 heavy
N_Geneseq_36:T31319	+ 100.50	139.69	3.52	1617	Antibody D heavy chain. Prodn.
N_Geneseq_36:T31320	+ 100.50	125.35	22.42	6284	Plasmid Hu19AHCpcd encoding an
N_Geneseq_36:T31321	+ 99.50	128.69	14.14	4157	Human CD100 antigen cDNA. Nucle
N_Geneseq_36:T31322	+ 99.50	124.82	23.70	6127	Mus musculus anti-IgE antibody
N_Geneseq_36:T31323	+ 99.00	143.04	2.29	914	Human EST R2810 DNA. Complexes
N_Geneseq_36:T31324	+ 99.00	135.50	5.72	1869	Human immunoglobulin heavy chain
N_Geneseq_36:T31325	+ 98.50	143.79	2.08	784	Anti-phenytoxazalone antibody c
N_Geneseq_36:T31326	+ 98.50	140.78	3.06	1060	Kidney injury associated molec
N_Geneseq_36:T31327	+ 97.00	135.75	5.83	1386	Anti-HIV-1 recombinant antibod

N_Geneseq_36:Q12632	+	97.00	130.42	11.55	2364	CD4-specific CDR-grafted he
N_Geneseq_36:Q12633	+	97.00	110.51	148.55	17350	hOP1 human osteogenic prot
N_Geneseq_36:Q53142	-	97.00	110.47	149.20	17410	Sequence encoding osteogen
N_Geneseq_36:T18381	-	97.00	110.47	149.20	17410	hOP-1 genomic DNA. Antibod
N_Geneseq_36:V15205	-	97.00	110.47	149.20	17410	Human osteogenic protein O

seq\_name: N\_Geneseq\_36:X28178

seq\_documentation\_block:

ID X28178 standard; cDNA; 2040 BP.  
AC X28178:  
DT 16-JUN-1999 (first entry)  
DE Human PIGRL-1 coding sequence.  
KW PIGRL-1; human; autoimmune disease; hyper-IgM Immunodeficiency; HIM;  
KW X-linked Severe Combined Immunodeficiency; XSCID; IgA deficiency;  
KW diagnosis; therapy; ss.  
OS Homo sapiens.  
PN EP-905238-A2.  
PD 31-MAR-1999. 306487.  
PF 14-AUG-1998; US-961564.  
PR 30-OCT-1997; US-961564.  
PR 25-AUG-1997; US-056935.  
PA (SMIK ) SMITHKLINE BEECHAM CORP.  
PI Sweet RW, Truneh A, Wu S;  
DR WPI: 99-192666/17.  
DR P-PSDB; Y05001.  
PT New polypeptides encoding PIGRL-1 useful for treating diseases such as X-linked Severe Combined Immunodeficiency  
PS Claim 4; Page 18-19; 26pp; English.  
CC This sequence encodes the human PIGRL-1 protein of the invention.  
CC Autoimmune diseases involving altered expression or activity of PIGRL-1 may include Hyper-IgM Immunodeficiency (HIM), X-linked Severe Combined Immunodeficiency (XSCID) and IgA deficiency. These diseases can be diagnosed or susceptibility to them predicted by: (1) determining whether there is a mutation in the genomic copy of the gene encoding PIGRL-1; or (2) measuring the amount of PIGRL-1 in a sample derived from the patient. Patients deficient in PIGRL-1 can be treated by administering either the PIGRL-1 DNA or its complement or an agonist of PIGRL-1 to the patient. Patients with excessive expression or activity of PIGRL-1 can be treated by administering an antagonist of PIGRL-1, an antisense nucleic acid molecule which inhibits the expression of PIGRL-1 or administering sufficient PIGRL-1 to compete with the endogenous activity. PIGRL-1 can be used to identify its agonists by contacting a cell expressing PIGRL-1 with a candidate compound in the presence of a signal system and noting the candidate as an agonist if a signal is produced. The same method can be used to identify antagonists of PIGRL-1 but the presence of an antagonist is indicated by a decrease in production of the signal. Antibodies against PIGRL-1 may be used to isolate or identify clones expressing PIGRL-1. Polynucleotides encoding PIGRL-1 may be used to identify chromosomal mutations in the gene encoding PIGRL-1 in patients. This information may then be correlated with the incidence of autoimmune disease in those patients to identify whether the mutation causes the disease.

alignment\_scores:  
Quality: 1340.00 Length: 255  
Ratio: 5.255 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-135-238B-2\_COPY\_18\_272 x X28178 ..  
Align seg 1/1 to: X28178 from: 1 to: 2040  
1 ArgileLeuProGluValValysValGluGlyGlyserValth 17  
|||||  
197 AGGATCTCCAGAGTAAGTAAGGAGGGGAGCTGGCGGATCAGTTAC 246  
|||||  
17 rileLysCysProLeuProGluMetHisValArgileTyrLeuCysArg 34  
|||||  
247 CATCAAGTGCCCACTTCTCTGAAATGCATGTGAGGATATATCTGTGCGGG 296

34 luMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsnPhe 50  
|||||  
297 AGATGGCTGGATCTGGAAATCATGTGTACCGTGGTATCCACCAACTTC 346  
51 IleLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProArgLy 67  
|||||  
347 ATCAGGCGAATATACAGGCGCGAGTTACTCTGAGCAATATCCAGCAA 396  
67 sAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerGlyV 84  
|||||  
397 GAATCTGTCTCTAGTGGAGGTACACAGCTCACAGAAAGTGACAGCGAG 446  
84 alTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLys 100  
|||||  
447 TCTATGCTTGGGAGCGGCGATGACACAGACCGGGGAAAGACCCAGAAA 496  
101 ValThrLeuAsnValHisSerGlyThrGluProSerTrpGluGlnPr 117  
|||||  
497 GTCAACCTGAATGTCCACAGTGAATACGAGCCATCATGGGAAGCAGCC 546  
117 oMetProGluThrProLysTrpPheHisLeuProTyrLeuPheGlnMetP 134  
|||||  
547 AATGCTGTGAGCTCAAAATGGTTTCATCTGCCCTATTGTTCCAGATGC 596  
134 roAlaTyrAlaSerSerSerLysPheValThrArgValThrProAla 150  
|||||  
597 CTGCATATGCCAGTTCTTCCAAATTCGTAACCCAGAGTTACACACAGCT 646  
151 GlnArgGlyLysValProProValHisHisSerSerProThrThrGlnI 167  
|||||  
647 CAAAGGGGCAAGGTCCTCCAGTTCACCACTCTCCCTCCCTCCACCCCAAT 696  
167 eThrHisArgProArgValSerArgAlaSerSerValAlaGlyAspLysP 184  
|||||  
697 CACCCACCCCTCCAGTGTCCAGAGCATCTTCAGTACGAGGTGACAAGC 746  
184 roArgThrPheLeuProSerThrThrAlaSerLysIleSerAlaLeuGlu 200  
|||||  
747 CCCGAACCTTCCTGCCATCCACTACAGCCTCAAAAATCTCAGCTCTGGAG 796  
201 GlyLeuLysProGlnThrProSerTyrAsnHisHisThrArgLeuHi 217  
|||||  
797 GGGTGTCTCAAGCCCGACCCAGCCCGAGCTACACACACACACAGGCTGCA 846  
217 sArgGlnArgAlaLeuAspTyrGlySerGlnSerGlyArgGluGlyGlnG 234  
|||||  
847 CAGGCAGAGACGACTGGACTATGGCTCACAGTCTGGGAGGAAGGCCAAG 896  
234 lypHeHisIleLeuIleProThrIleLeuGlyLeuPheLeuAlaLeu 250  
|||||  
897 GATTTCATCTGATCTGATCCCGACCATCTGGGCTTTTCTCTCTGGCATT 946  
251 LeuGlyLeuValVal 255  
|||||  
947 CTGGGGCTGGTGGTG 961  
seq\_name: N\_Geneseq\_36.X28179  
seq\_documentation\_block:  
ID X28179 standard; cDNA; 1047 BP.  
AC X28179;  
DT 16-JUN-1999 (first entry)  
DE Human PIGRL-1 coding sequence fragment.  
KW PIGRL-1; human; autoimmune disease; hyper-IgM Immunodeficiency; HIM;  
KW x-linked Severe Combined Immunodeficiency; XSCID; IgA deficiency;  
KW diagnosis; therapy; ss.  
OS Homo sapiens.  
PN EP-905238-A2.  
PD 31-MAR-1999.  
PF 14-AUG-1998; 306487.  
PR 30-OCT-1997; US-961564.  
PR 25-AUG-1997; US-056935.

PA (SMIK ) SMITHKLINE BEECHAM CORP.  
PI Sweet RW, Truneh A, Wu S;  
DR WPI: 99-192666/17.  
P-PSDB: Y05002.  
PT New polypeptides encoding PIGRL-1 useful for treating diseases such  
as x-linked Severe Combined Immunodeficiency  
PS Disclosure: Page 8; 26pp; English.  
CC This sequence encodes the human PIGRL-1 protein of the invention.  
CC Autoimmune diseases involving altered expression or activity of PIGRL-1  
CC may include Hyper-IgM Immunodeficiency (HIM), x-linked Severe Combined  
CC Immunodeficiency (XSCID) and IgA deficiency. These diseases can be  
CC diagnosed or susceptibility to them predicted by: (1) determining whether  
CC there is a mutation in the genomic copy of the gene encoding PIGRL-1; or  
CC (2) measuring the amount of PIGRL-1 in a sample derived from the patient.  
CC Patients deficient in PIGRL-1 can be treated by administering either the  
CC PIGRL-1 DNA or its complement or an agonist of PIGRL-1 to the patient.  
CC Patients with excessive expression or activity of PIGRL-1 can be treated  
CC by administering an antagonist of PIGRL-1, an antisenase nucleic acid  
CC molecule which inhibits the expression of PIGRL-1 or administering  
CC sufficient PIGRL-1 to compete with the endogenous activity. PIGRL-1 can  
CC be used to identify its agonists by contacting a cell expressing PIGRL-1  
CC with a candidate compound in the presence of a signal system and noting  
CC the candidate as an agonist if a signal is produced. The same method can  
CC be used to identify antagonists of PIGRL-1 but the presence of an  
CC antagonist is indicated by a decrease in production of the signal.  
CC Antibodies against PIGRL-1 may be used to isolate or identify clones  
CC expressing PIGRL-1. Polynucleotides encoding PIGRL-1 may be used to  
CC identify chromosomal mutations in the gene encoding PIGRL-1 in patients.  
CC This information may then be correlated with the incidence of autoimmune  
CC disease in those patients to identify whether the mutation causes the  
CC disease.  
SQ Sequence 1047 BP; 255 A; 281 C; 276 G; 223 T;  
alignment\_scores:  
Quality: 1034.00 Length: 258  
Ratio: 4.345 Gaps: 12  
Percent Similarity: 92.248 Percent Identity: 89.922  
alignment\_block:  
US-09-135-238b-2\_copy\_18\_272 x X28179 ..  
Align seg 1/1 to: X28179 from: 1 to: 1047  
1 ArgIleLeuProGluValLysValGluGlyLeuGlyGlySerValTh 17  
|||||  
201 AGGATCTCCAGAGCTAAAGGTAGAGGGGAGCTGGCGGATCAGTTAC 250  
17 rIleLysCysProLeuProGluMetHisValArgIleTyrLeuCysArgG 34  
|||||  
251 CATCAAGTGCCTCTCTGAAATGCATGTGAGGATATATCTGTCCCGGG 300  
34 luMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsnPhe 50  
|||||  
301 AGATGGCTGGATCTGGAGCATGTGTACCGTGGTATCCACCAACTTC 350  
51 IleLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProArgLy 67  
|||||  
351 ATCAAGGCGAATATACAGGCGCGAGTTACTCTGAAGCAATATCCACGCAA 400  
67 sAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerGlyV 84  
|||||  
401 GAATCTGTCTCTAGTGGAGGTACACAGCTGACAGAAAGTGACAGCGAG 450  
84 alTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLys 100  
|||||  
451 TCTATGCTTGGGAGCGGCGATGACACAGACCGGGGAAAGACCCAGAAA 499  
101 ValThrLeuAsnValHisSerGlyThrGluProSerTrpGluGlnPr 117  
|||||  
500 GTCACCCCTGAATGTCCACAGTGAATACGAGCCATCATGGGAAGCAGGCC 549  
117 oMetProGluThrProLysTrpPheHisLeuProTyrLeuPheGlnMetP 134



```

|||||
550 AATGCTCAGACTCCAAATGGTTTCATCTGCCCTATTTGTCAGATGC 599
134 toAlaTyrAlaSerSerSerLysPheValThr..ArgValThrThrProA 150
|||||
600 CTGCATATGCCGGTCTTCCACATTCGTAACCGCAGAGATTACCACACAG 649
150 la.GlnArgGlyLysValProValHisHisSerSerProThrThrG1 166
650 CTTCAAGGGGCAAGGTCCTCCAGTTCACCACTCTCCCTCCACCACCCA 699
166 nile..ThrHisArgPro.ArgValSerArgAlaSerSerValAlaGlyAs 182
700 AATTCACCCACCGCCCTTCGAGTTCGAGTTCGAGCATCTTCAGTAGCAGGTGA 749
182 pLysProArg..ThrPheLeuProSerThrThrAlaSerLysIleSerAla 198
750 CAAGCCCCGAAACITTCCTGCCATCCACTACAGGCTCAAAAATCTCAGCT 799
199 LeuGluGlyLeuLeuLys....ProGlnThrProSerTyrAsnHisHisTh 214
800 CTGGAAGGGCTGCTTCAAGCCCCAGAGAGGCCAGCTACAAACAN.CACAC 848
214 tArgLeuHisArgGlnArgAlaLeuAsp...Tyr.GlySerGlnSerGly 229
849 CAGGCTGCACAGGACAGAGACTGGATCTATGCGGNTCACAGTCTGGG 898
230 ArgGluGly.....GlnGlyPhe...HisIleLeuIlePro..ThrIleL 243
899 ...GAGGGGAANGNCAGAGGATTTTACATTCCTTGATTCGCGGACCATCN 945
243 euGlyLeuPhe 246
||| |||
946 FTGGGGCCCTTT 956

```

seq\_name: N\_Geneseq\_36:T31290

seq\_documentation\_block:

```

ID T31290 standard; cdNA; 3095 BP.
AC T31290;
DE 24-FEB-1997 (first entry)
DE Mouse poly-immunoglobulin receptor, cdNA.
KW Mouse; immunoglobulin; receptor; protection protein; mutants;
KW heavy chain; antigen binding domain; protection; pathogen;
KW mucosal; environment; gastrointestinal; passive; immunisation;
KW Guy's 13 antibody; prevention; dental caries; Streptococcus;
KW poly; sorbinus; murine; ss.
OS Mus musculus.
PE Key Location/Qualifiers
FT cds 85..2400
FT /*tag= a
FN W09621012-A1.
FD 11-JUL-1996.
PF 27-DEC-1995; U16889.
PR 30-DEC-1994; US-367395.
PR 04-MAY-1995; US-434000.
PR (PLAN-) PLANT BIOTECHNOLOGY INC..
PR (UNWE-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.
PR (PLAN-) PLANET BIOTECHNOLOGY INC.
PI Hiatt AC, Lehner T, Ma JKC;
DR WPI; 96-333987/33.
DR P-PSDB: W03180.
PT Immunoglobulin and protection protein complex and its prodn. in
PT plants - useful for passive immunisation against mucosal antigens,
PT esp. against S. mutans and S. sorbinus to prevent dental caries
PS Disclosure; pages 117-121; 152pp; English.
CC The present sequence encodes the mouse poly-immunoglobulin (Ig)
CC receptor, a portion of which corresp. to residues 1-627, pref.
CC 1-606, or esp. residues 13-45, 1-120, 110-230, 210-340, 320-450.
CC 440-550, 550-605 or 550-627 comprises a protection protein (PP)..
CC The Ig of the invention comprises a pp as above in association with
CC an Ig derived heavy chain, having at least a portion of an antigen
CC (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.

```

```

CC gastrointestinal, environments, therefore enhancing its
CC effectiveness in passively immunising animals against mucosal
CC pathogens. The Ag binding domain is specifically derived from the
CC Guy's 13 antibody, and the Ig can be used to prevent dental caries
CC by binding, e.g. Streptococcus mutans serotypes c, e and f, or
CC S. sorbinus serotypes d and g.
SQ Sequence 3095 BP; 861 A; 796 C; 784 G; 654 T;

```

alignment\_scores:

```

Quality: 197.00 Length: 148
Ratio: 2.165 Gaps: 5
Percent Similarity: 61.486 Percent Identity: 32.432

```

alignment\_block:

US-09-135-238B-2\_copy\_18\_272 x T31290 ..

Align seg 1/1 to: T31290 from: 1 to: 3095

```

13 GlyGlySerValThrIleLysCysProLeuProGlu.....Me 25
||| ||||| ||||| ||||| |||||
181 GGCACACTCTGTTCCATCAGCTGCTACTACCCAGACACCTCTGTCAACCG 230
25 tHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 42
||| ||||| ||| ||||| ||| ||||| |||
231 GCACACCCGGGAATACTGTTGCCGACACAGGAGCC...AGCGGCATGTGCA 277
42 lyThrValValSerThrThrAsnPheIleLysAlaGluTyrLysGlyArg 58
||||| ||||| ||||| ||||| ||||| ||||| |||||
278 CAAGGCTCATCTCTCAATGGGTACCTCTCCAAGGAGTATTCCAGGCAGA 327
59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75
||||| ||||| ||||| ||||| ||||| ||||| |||||
328 GCCAACCTCATCAACTCCAGAGACACACACATTTGTATTACCATGA 377
75 rGlnLeuThrGluSerSerGlyValTyrAlaCysGlyAlaGlyMetA 92
||||| ||||| ||||| ||||| ||||| ||||| |||||
378 GCAGCTCACCAGCAGCAGCAGCTGGGAGCTACAAGTGTGGCTGGGT...A 424
92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 108
425 CCAGTACCCGAGGCTGCTCTTCGATCTCAGCTGGAGTGC..... 465
109 TyrGluProSerTyrGluGlnProMetProGluThrProLysTrpPh 125
||| ||||| ||||| ||||| ||||| ||||| |||||
466 .....AGCCAGGTTCTCTGAGTTGCCGAGTGACAC 494
125 eHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSerSerSerLysP 142
||||| ||||| ||||| ||||| ||||| ||||| |||||
495 CCAGGTC.....TACACAAGGACATAGGCA 520
142 heValThrArgValThrThrProAlaGlnArgGlyLysValPro 156
||| ||| ||| ||| ||| ||| ||| ||| |||
521 GAAATGTGACCATTTGAATGCCITTCAAAAGGGAGAATGTTCC 564

```

seq\_name: N\_Geneseq\_36:V15422

seq\_documentation\_block:

```

ID V15422 standard; DNA; 29392 BP.
AC V15422;
DE 11-JUN-1998 (first entry)
DE Mouse poly Ig receptor protein gene.
KW Mouse; poly Ig receptor protein; pIgR protein; pIg; deficiency;
KW knockout mouse; disease model; ds.
OS Mus sp.
PN J10057066-A.
PD 03-MAR-1998.
PF 19-AUG-1996; 217154.
PR 19-AUG-1996; JP-217154.
PR (HONS) YAKULT HONSHA KK.
DR WPI; 98-254323/23.
PT Mouse pIg receptor protein gene - used for preparing gene knockout
PT mice, useful for study of human poly Ig receptor protein deficiency

```

PS Claim 1; Page 4-14; 18pp; Japanese.

CC The present sequence represents the mouse poly Ig receptor protein gene, which has a 29392 bp sequence. The new gene can be used to produce a gene knockout mouse, useful as a disease model of human poly Ig receptor protein deficiency.

SQ Sequence 29392 BP; 8318 A; 6747 C; 6514 G; 7813 T;

#### alignment\_scores:

Quality: 188.00 Length: 97  
Ratio: 2.648 Gaps: 3  
Percent Similarity: 73.196 Percent Identity: 40.206

#### alignment\_block:

US-09-135-238B-2\_COPY\_18\_272 x V15422 ..

Align seg 1/1 to: V15422 from: 1 to: 29392

13 GlyGlySerValThrIleLysCysProLeuProGlu.....Me 25  
||||| ||||| ||||| ||||| |||||  
18245 GCGGACCTGTTCCATCAGCTGCTACTACCCAGACACCTCTGTCAACCG 18294  
  
25 thisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 42  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
18295 GCACACCCGGAATACTGTGTCGCCAGAGAGCC...AGCGGATGTGCA 18341  
  
42 LyThrValValSerThrThrAsnPheIleLysAlaGluTyrLysGlyArg 58  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
18342 CAACGCTCATCTTCCAAATGGCTACCTCTCCAGAGATATTCAGGCAGA 18391  
  
59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
18392 GCCAAGCTCATCACTTCCAGAGAGACACACATTTGTGATTACATTGA 18441  
  
75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMetA 92  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
18442 GCAGCTCACCCAGGAGGACCTGGAGCTACAAGTGTGGCTGGGT...A 18488  
  
92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnVal 105  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
18489 CCAGTACCCAGGAGGCTGCTTCGATGTCAGCTGGAGGTC 18529

seq\_name: N\_Geneseq\_36:T31291

#### seq\_documentation\_block:

ID T31291 standard; CDNA; 3269 BP.  
AC T31291; 1997 (first entry)  
DT 24-FEB-1997 (first entry)  
DE Rat poly-immunoglobulin receptor, cDNA.  
KW Rat; immunoglobulin; receptor; protection protein; mutants;  
heavy chain; antigen binding domain; protection; pathogen;  
mucosal; environment; gastrointestinal; passive; immunisation;  
KW Guy's 13 antibody; prevention; dental caries; Streptococcus;  
KW poly; sorbinus; ss.  
OS Rattus rattus.  
FH Key Location/Qualifiers  
FT cds 74..2383 /\*tag= a  
FT WO9621012-AL.  
PN 11-JUL-1996.  
PF 27-DEC-1995; U16889.  
PR 30-DEC-1994; US-367395.  
PR 04-MAY-1995; US-434000.  
PA (PLAN-) PLANT BIOTECHNOLOGY INC.  
PA (UNWE-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.  
PA (PLAN-) PLANET BIOTECHNOLOGY INC.  
PI Hiatt AC, Lehner T, Ma JKC;  
DR WPI; 96-333987/33.  
DR P-PSDB; W03181.  
PT Immunoglobulin and protection protein complex and its prodn. in  
plants - useful for passive immunisation against mucosal antigens,  
esp. against S. mutans and S. sorbinus to prevent dental caries  
PT Disclosure; Pages 123-127; 152pp; English.

CC The present sequence encodes the rat poly-immunoglobulin (Ig) receptor, a portion of which corresp. to residues 1-627, pref. 1-606, or esp. residues 13-45, 1-120, 110-230, 210-340, 320-450, 440-550, 550-606 or 550-627 comprises a protection protein (pp).  
CC The Ig of the invention comprises a pp as above in association with an Ig derived heavy chain, having at least a portion of an antigen (Ag) binding domain. The pp protects the Ig in harsh mucosal, e.g. gastrointestinal, environments, therefore enhancing its effectiveness in passively immunising animals against mucosal pathogens. The Ag binding domain is specifically derived from the Guy's 13 antibody, and the Ig can be used to prevent dental caries by binding, e.g. Streptococcus mutans serotypes c, e and f, or S. sorbinus serotypes d and g.  
SQ Sequence 3269 BP; 890 A; 831 C; 850 G; 698 T;

#### alignment\_scores:

Quality: 185.50 Length: 164  
Ratio: 1.855 Gaps: 6  
Percent Similarity: 60.976 Percent Identity: 30.488

#### alignment\_block:

US-09-135-238B-2\_COPY\_18\_272 x T31291 ..

Align seg 1/1 to: T31291 from: 1 to: 3269

13 GlyGlySerValThrIleLysCysProLeuProGlu.....Me 25  
||||| ||||| ||||| ||||| |||||  
170 GGTAACTCGTCTCCATCAGCTGCTACTACCCAGACACCTCTGTCAACCG 219  
  
25 thisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 42  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
220 GCACACCCGGAATACTGTGTCGCCAGAGAGCC...AACGGTACTGGG 266  
  
42 LyThrValValSerThrThrAsnPheIleLysAlaGluTyrLysGlyArg 58  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
267 CAACGCTCATCTTCCAAATGGCTACCTCTCGAAGAGATATTCAGGCAGA 316  
  
59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
317 GCAGCTCATCACTTCCAGAGAGACATTTGTGATTACATTGC 366  
  
75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMetA 92  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
367 ACATCTCACCCAGGAGGACACTGGAGCTACAAGTGTGGTGGGT...A 413  
  
92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 108  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
414 CCATATACCCAGGAGGCTGTTTTCGATGTCAGCTGGAGGTC..... 454  
  
109 TyrGluProSerTrpGluGlnProMetProGluThrProLysTrpPh 125  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
455 .....AGCCAGGTTCTGAGTCCCAATGACAC 483  
  
125 eHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSerSerSerLysP 142  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
484 CCATGTC.....TACACAAAGG 500  
  
142 heValThrArgValThrThrProAlaGlnArgGlyLysValProProVal 158  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
501 ACATAGGCAGACACTGTGACCATCGAATGCCGTTTCAAGAGGGGAATGCT 550  
  
159 HisHisSer\_SerProThrThrGlnIleThrHisArgPro 171  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
551 CATAGCAAGAAATCCCTGTGTAAGAGAGAGAGAGGCGCT 590

seq\_name: N\_Geneseq\_36:X07407

#### seq\_documentation\_block:

ID X07407 standard; CDNA to mRNA; 1839 BP.  
AC X07407;  
DT 08-JUN-1999 (first entry)  
DE Human secretory immunoglobulin A component gene.



[illegible]

seq\_name: N\_Geneseq\_36:T30856

```

seq_documentation_block:
ID  T30856 standard; cDNA; 2031 BP.
AC  T30856;
DE  13-SEP-1996 (first entry)
DE  Secretory component DNA fragment 1.
DE  Secretory component; SC DNA fragment 1; crystallisation;
KW  PolymERIC Immunoglobulin receptor; pIgR; CHO; ss.
OS  Chimeric Homo sapiens;
OS  Chimeric synthetic.
FH  Key Location/Qualifiers
FT  misc_feature 1..33
FT  /tag= a
FT  /note= "bases 1-33 derived from PCR with 5'
FT  primer 1"
FT  111..2015
FT  /tag= b
FT  /product= transcript from SC DNA fragment 1
FT  complement (1997. :2031)
FT  /tag= c
FT  /note= "from PCR with 3' primer"
FT  WO9618734-AL.
PN  20-JUN-1996.
PD  PD
PF  06-DEC-1995; E04797.
PR  16-DEC-1994; EP-120019.
PR  (CIBA ) CIBA GEIGY AG.
PA  Antibergs FAM, Gruetter M, Leist C. Van Oostrum J;
PI  WPI: 96-300651/30.
PR  prodn. of recombinant secretory protein capable of crystallisation -
PR  pref. human pIgR, for use in screening and binding studies
PR  Claim 5: Page 36-38; 50pp; English.
CC  SC DNA fragment 1 (T30856) was obtd. by PCR amplification of human
CC  polymERIC immunoglobulin receptor (pIgR) cDNA in vector pcH6 using a
CC  5' primer (T30859) hybridising to the cytomegalovirus promoter in
CC  pcH6 and a 3' primer (T30861) that generates a stop codon at
CC  position 1906 just upstream of the hydrophobic transmembrane segment
CC  of the pIgR and also creates a downstream XbaI site. CHO SFR 3
CC  cells transfected with a vector carrying the amplified DNA can be
CC  used for prodn. of recombinant secretory component (SC). This can
CC  be crystallised for receptor structure studies, used to stabilise
CC  immunoglobulins, or used to screen (ant)agonists capable of
CC  modulating mucosal immune responses.
SQ  sequence 2031 BP: 477 A; 605 G; 408 T;

```

4	ProGluValIysValGluGlyGluLeuGlyGlySerValThrIleLysCys	20
183	CCCCGAG...GAGGTGAATGTGGAAAGGTAACTCACTGCTCCATCAGGTG	229
20	sProLeuPro.....GluMetHisValArgIleTyrLeuCysAs	33
230	CTACTACCCACCCACTCTGTCAACCGGCACACCGGAAAGTACTGCTGCC	279
33	rgGluMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsn	49
280	GGCAGGAGAGCTAGAGGTGGC...TGCATAAACCTTCATCTCTCGGAGGCG	326
50	PheIleLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProAr	66
327	TACGTCTCCAGCAAAATATGCAGCGAGGCGTAACCTCACCAAGCTCCCGGA	376
66	gLyAspAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerG	83
377	GAAAGGCACAAATCGTGTGTGAACATTCGCCAGCTGAGCGAGGATGATCCG	426
83	lyValTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGln	99
427	GGCCTCAAGTGTGGCTGGGATCAATAGC...CGAGGCGCTGTCCCTTT	473
100	LysValThrLeuAsnVal	105
474	GATGTCAGCTGGAGGTC	491
seq_name: N_Geneseq_36:T31288		
seq_documentation_block:		
ID	T31288 standard; cDNA; 2919 BP.	
AC	T31288;	
DT	24-FEB-1997 (first entry)	
DE	Human poly-immunoglobulin receptor, cDNA.	
DE	Human; immunoglobulin; receptor; protection protein; mutans;	
KW	heavy chain; antigen binding domain; protection; pathogen;	
KW	mucosal; environment; gastrointestinal; passive; immunisation;	
KW	Guy's 13 antibody; prevention; dental caries; Streptococcus;	
KW	poly; sorbinus; ss.	
OS	Homo sapiens.	
FT	Location/Qualifiers	
FT	key	
FT	235..2475	
FT	/*tag= a	
PT	W09621012-11.	
PD	11-JUL-1995.	
PD	27-DEC-1995; U16889.	
PF	30-DEC-1994; US-367395.	
PF	04-MAY-1995; US-434000.	
PR	(PLAN-) PLANT BIOTECHNOLOGY INC.	
PA	(UNNE-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.	
PA	(PLAN-) PLANET BIOTECHNOLOGY INC.	
PI	Hiatt AC, Lehner T, Ma JKC;	
PI	WPI: 96-333987/33.	
DR	P-PSDB: W03178.	
DR	Immunoglobulin and protection protein complex and its prodn. in	
PT	plants - useful for passive immunisation against mucosal antigens,	
PT	esp. against S. mutans and S. sorbinus to prevent dental caries	
PS	Disclosure: Pages 105-108; 152pp; English.	
PS	The present sequence encodes the human poly-immunoglobulin (Ig)	
CC	receptor, a portion of which corresp. to residues 1-627, pref.	
CC	1-606, or esp. residues 13-45, 1-120, 110-230, 210-340, 320-450,	
CC	440-550, 550-606 or 550-627 comprises a protection protein (PP).	
CC	The Ig of the invention comprises a PP as above in association with	
CC	an Ig derived heavy chain, having at least a portion of an antigen	
CC	(Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.	
CC	gastrointestinal, environments, therefore enhancing its	
CC	effectiveness in passively immunising animals against mucosal	
CC	pathogens. The Ag binding domain is specifically derived from the	
CC	Guy's 13 antibody, and the Ig can be used to prevent dental caries	
CC	by binding, e.g. Streptococcus mutans serotypes c, e and f, or	
CC	S. sorbinus serotypes d and g.	

FD  
I / - MAR - 1998 .

```

418 GAGCGGCACGTTTGTGGTGCATCAGCCCATCCACCATAAAGACTCAG 457
      ||| ||||| |||:::|||||
83 lyValTyraLacCysGlyAlaIagLyMetAsnThrAspargGlyLysThrGln 99
      || ||| ||||| |||:::|||||
468 GCGCTACAAGTGTGCCCTGGGCATTAGCAGC...CGTGGCCTTAAC TTC 514
      |||:::||||| ::::
100 LysValThrLeuAsnValHisSerGlu 108
      |||:::||||| ::::
515 GATGTGAGCCTGGAGGTGAGCCAAGAT 541

seq_name= N_Geneseq_36.V20383

seq_documentation_block:
ID V20383 standard; CDNA to mRNA; 1911 BP.
AC V20383;
DE 26-JUN-1998 (first entry)
DT CDNA for human immunity related factor.
KW lymph node; human; immunity related factor; research; treatment;
KW immune disease; infectious disease; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS           98..1696
                  /*tag= a
FT sig_peptide   98..145
                  /*tag= b
FT mat_peptide   146..1693
                  /*tag= c
                  /product= immunity_related_factor
FT J10072495-A.
PN 17-MAR-1998.
PD
```



Align seg 1/1 to: Q14498 from: 1 to: 1876

```

13 GlyGlySerValThrIleLysCysProLeuPro.....GluMe 25
|||||
238 GGGGACTCGGTGTCATCATCTACTACCCCAACACCTCGTCACCGC 287
25 tHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 42
|||||
288 GCACAGCCGGAAGTTCTGTCGCGGAA...GAGGAGAGCGCGCTGCG 334
42 lyThrValValSerThrThrAsnPhelLysAlaGluTyrLysGlyArg 58
|||||
335 TGACGCTTGCTCG...ACGGGTACACGTCCAGGAATACTCCGGGAGA 381
59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75
|||||
382 GGCAAGCTCAGGACTTCCTCTGATAAGGGGAGTTTGTGGTACTGTTGA 431
75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 92
|||||
432 CCAACTCACCCAGACGACTCAGGAGCTACAAGTGTGCGTGGGAGTCA 481
92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 108
|||||
482 AC...GGCGGTGGCTGGACTTCGCTGTCAACGTGCTGTCAGCCAGAG 528
109 TyrGluPro 111
|||||
529 CCAGAGCCT 537

```

seq\_name: N\_Geneseq\_36:T31287

seq\_documentation\_block:

ID T31287 standard; cDNA; 3517 BP.

AC T31287; 1997 (first entry)  
 DT Rabbit poly-immunoglobulin receptor, cDNA.  
 DE Rabbit poly-immunoglobulin receptor; protection protein; mutants;  
 KW heavy chain; antigen binding domain; protection; pathogen;  
 KW mucosal; environment; gastrointestinal; passive; immunisation;  
 KW Guy's 13 antibody; prevention; dental caries; Streptococcus;  
 KW poly; sorbinus; ss.  
 OS Oryctolagus cuniculus.  
 FH Key Location/Qualifiers  
 FT 124..2445  
 FT cds /\*tag= a  
 PN W09621012-A1.  
 PD 11-JUL-1996.  
 PF 27-DEC-1995; U16889.  
 PR 30-DEC-1994; US-367395.  
 PR 04-MAY-1995; US-434000.  
 PA (PLAN-) PLANT BIOTECHNOLOGY INC.  
 PA (UNME-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.  
 PI Hiatt AC, Lehner T, Ma JKC.  
 DR WPI; 96-333987/33.  
 DR P-PSDB; W03177.  
 PT Immunoglobulin and protection protein complex and its prodn. in  
 PT plants - useful for passive immunisation against mucosal antigens,  
 PT esp. against S. mutans and S. sorbinus to prevent dental caries  
 PS Claim 10; Pages 99-102; 152pp; English.  
 CC The present sequence encodes the rabbit poly-immunoglobulin (Ig)  
 CC receptor, a portion of which corres. to residues 1-627, pref.  
 CC 1-606, or esp. residues 21-43, 1-118, 119-223, 224-332, 333-441,  
 CC 442-552, 553-606 or 532-627 comprises a protection protein (PP).  
 CC The Ig of the invention comprises a PP as above in association with  
 CC an Ig derived heavy chain, having at least a portion of an antigen  
 CC (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.  
 CC gastrointestinal, environments, therefore enhancing its  
 CC effectiveness in passively immunising animals against mucosal  
 CC pathogens. The Ag binding domain is specifically derived from the  
 CC Guy's 13 antibody, and the Ig can be used to prevent dental caries

CC by binding, e.g. Streptococcus mutans serotypes c, e and f, or  
 CC S. sorbinus serotypes d and g.  
 SQ Sequence 3517 BP; 795 A; 1027 C; 1038 G; 657 T;

alignment\_scores:

Quality: 166.50 Length: 103  
 Ratio: 2.602 Gaps: 4  
 Percent Similarity: 62.136 Percent Identity: 42.718

alignment\_block:

US-09-135-238B-2\_COPY\_18\_272 x T31287 ..

Align seg 1/1 to: T31287 from: 1 to: 3517

```

13 GlyGlySerValThrIleLysCysProLeuPro.....GluMe 25
|||||
238 GGGGACTCGGTGTCATCATCTACTACCCCAACACCTCGTCACCGC 287
25 tHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 42
|||||
288 GCACAGCCGGAAGTTCTGTCGCGGAA...GAGGAGAGCGCGCTGCG 334
42 lyThrValValSerThrThrAsnPhelLysAlaGluTyrLysGlyArg 58
|||||
335 TGACGCTTGCTCG...ACGGGTACACGTCCAGGAATACTCCGGGAGA 381
59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75
|||||
382 GGCAAGCTCAGGACTTCCTCTGATAAGGGGAGTTTGTGGTACTGTTGA 431
75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 92
|||||
432 CCAACTCACCCAGACGACTCAGGAGCTACAAGTGTGCGTGGGAGTCA 481
92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 108
|||||
482 AC...GGCGGTGGCTGGACTTCGCTGTCAACGTGCTGTCAGCCAGAG 528
109 TyrGluPro 111
|||||
529 CCAGAGCCT 537

```

seq\_name: N\_Geneseq\_36:Q14499

seq\_documentation\_block:

ID Q14499 standard; cDNA; 1876 BP.  
 AC Q14499;  
 DT 30-JAN-1992 (first entry)  
 DE Clone 2 for truncated poly Ig-receptor.  
 KW Rabbit; insemination; pregnancy; ss.  
 OS Oryctolagus cuniculus.  
 FH Key Location/Qualifiers  
 FT cds 124..1876  
 PN W09116061-A.  
 PD 31-OCT-1991.  
 PF 16-APR-1991; U02604.  
 PR 16-APR-1990; US-510161.  
 PA (HARD ) HARVARD COLLEGE.  
 PA (SURE-) INST SUISSE RECH EXPER C.  
 PI Kraehenbuhl JP, Weltzin RA, Neutra MR;  
 DR WPI; 91-339549/46.  
 DR P-PSDB; R14671.  
 PT Stabilised poly-Ig complex contg. portion of poly-Ig receptor -  
 PT useful in protection against pathogens or against pregnancy  
 PS Disclosure; Fig 8; 51 pp; English.  
 CC The sequence was obtd. from a cDNA clone of allele no. 2 and  
 CC encodes a truncated poly-Ig receptor. The native gene (Mostov et  
 CC al) is mutated to delete the portion encoding the transmembrane and  
 CC intra-cellular domains. The recombinant protein produced by  
 CC expression of the sequence is used as a stabiliser protein with a  
 CC poly-Ig specific for a selected antigen or family of antigens. The  
 CC compsn. can be administered directly to the mucosal surfaces of a









GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 12, 2000, 10:09:19 ; Search time 3928.49 Seconds  
(without alignments)  
2144.023 Million cell updates/sec

Title: US-09-135-238B-1

Perfect score: 1910

Sequence: 1 aaaggagtaagcagcggtgc.....ttactctgtctcctcttt 1910

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 5247842 seqs, 2204914090 residues

Total number of hits satisfying chosen parameters: 10495684

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

- 1: em\_est1:\*
- 2: em\_est2:\*
- 3: em\_est3:\*
- 4: em\_est4:\*
- 5: em\_est5:\*
- 6: em\_est6:\*
- 7: em\_est7:\*
- 8: em\_est8:\*
- 9: em\_est9:\*
- 10: em\_est10:\*
- 11: em\_est11:\*
- 12: em\_est12:\*
- 13: em\_est13:\*
- 14: em\_est14:\*
- 15: em\_est15:\*
- 16: em\_est16:\*
- 17: em\_est17:\*
- 18: em\_est18:\*
- 19: em\_est19:\*
- 20: gb\_est1:\*
- 21: gb\_est2:\*
- 22: gb\_est3:\*
- 23: gb\_est4:\*
- 24: gb\_est5:\*
- 25: gb\_est6:\*
- 26: gb\_est7:\*
- 27: gb\_est8:\*
- 28: gb\_est9:\*
- 29: gb\_est10:\*
- 30: gb\_est11:\*
- 31: gb\_est12:\*
- 32: gb\_est13:\*
- 33: gb\_est14:\*
- 34: gb\_est15:\*
- 35: gb\_est16:\*
- 36: gb\_est17:\*
- 37: gb\_est18:\*
- 38: gb\_est19:\*
- 39: gb\_est20:\*
- 40: gb\_est21:\*
- 41: gb\_est22:\*
- 42: gb\_est23:\*
- 43: gb\_est24:\*

- 44: gb\_est25:\*
- 45: em\_est20:\*
- 46: em\_est21:\*
- 47: em\_est22:\*
- 48: em\_est23:\*
- 49: em\_est24:\*
- 50: em\_est25:\*
- 51: em\_est26:\*
- 52: em\_est27:\*
- 53: em\_est28:\*
- 54: em\_est29:\*
- 55: em\_est30:\*
- 56: em\_est31:\*
- 57: em\_est32:\*
- 58: em\_est33:\*
- 59: em\_est34:\*
- 60: em\_est35:\*
- 61: em\_est36:\*
- 62: em\_est37:\*
- 63: gb\_est26:\*
- 64: gb\_est27:\*
- 65: gb\_est28:\*
- 66: gb\_est29:\*
- 67: gb\_est30:\*
- 68: gb\_est31:\*
- 69: gb\_est32:\*
- 70: gb\_est33:\*
- 71: gb\_est34:\*
- 72: gb\_est35:\*
- 73: gb\_est36:\*
- 74: gb\_est37:\*
- 75: gb\_est38:\*
- 76: gb\_est39:\*
- 77: gb\_est40:\*
- 78: gb\_est41:\*
- 79: gb\_est42:\*
- 80: gb\_est43:\*
- 81: gb\_est44:\*
- 82: gb\_est45:\*
- 83: gb\_est46:\*
- 84: gb\_est47:\*
- 85: gb\_est48:\*
- 86: gb\_est49:\*
- 87: gb\_est50:\*
- 88: gb\_est51:\*
- 89: gb\_est52:\*
- 90: gb\_est53:\*
- 91: gb\_est54:\*
- 92: gb\_est55:\*
- 93: gb\_gss1:\*
- 94: gb\_gss2:\*
- 95: gb\_gss3:\*
- 96: gb\_gss4:\*
- 97: em\_gss1:\*
- 98: em\_gss2:\*
- 99: em\_gss3:\*
- 100: em\_gss4:\*
- 101: gb\_gss15:\*
- 102: gb\_gss16:\*
- 103: gb\_gss17:\*
- 104: gb\_gss18:\*
- 105: gb\_gss19:\*
- 106: em\_gss5:\*
- 107: em\_gss6:\*
- 108: em\_gss7:\*
- 109: em\_gss8:\*
- 110: em\_gss9:\*
- 111: em\_gss10:\*
- 112: em\_gss11:\*
- 113: em\_gss12:\*
- 114: em\_gss13:\*
- 115: gb\_gss5:\*
- 116: gb\_gss6:\*

117: gb\_gss7:\*  
118: gb\_gss8:\*  
119: gb\_gss9:\*  
120: gb\_gss10:\*  
121: gb\_gss11:\*  
122: gb\_gss12:\*  
123: gb\_gss13:\*  
124: gb\_gss14:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	635.8	33.3	711	44	AI760373	AI760373 wh87d11.x
C 2	516.4	27.0	578	35	AI084226	AI084226 oy72g09.x
C 3	458.8	24.0	524	79	AW452045	AW452045 UI-H-B13-
C 4	431	22.6	509	44	AI760350	AI760350 wh87b03.x
C 5	431	22.6	511	44	AI762709	AI762709 wh93d08.x
C 6	418.2	21.9	456	39	AI399732	AI399732 th28d08.x
C 7	418	21.9	488	30	AA766229	AA766229 oa28d09.s
C 8	409.2	21.4	453	30	AA769441	AA769441 nz41f03.s
C 9	403.2	21.1	441	73	AW001912	AW001912 wt81f03.x
C 10	400	20.9	458	44	AI761021	AI761021 w10b11.x
C 11	399	20.9	476	26	AA456309	AA456309 aal3g04.s
C 12	390.8	20.5	474	30	AA703531	AA703531 zj13b01.s
C 13	382.4	20.0	410	79	AA401870	AA401870 UI-HF-BK0
C 14	380	19.9	404	26	AA456707	AA456707 aal3g04.f
C 15	374	19.6	528	90	W01276	W01276 za40f12.r1
C 16	360.2	18.9	418	40	AI475802	AI475802 tc94f09.x
C 17	359	18.8	444	35	AI141006	AI141006 oy68f06.x
C 18	357.4	18.7	411	30	AA744586	AA744586 ny25g04.s
C 19	353	18.5	411	31	AA806272	AA806272 oc26a03.s
C 20	348.8	18.3	390	27	AA502725	AA502725 ng63d12.s
C 21	337	17.6	416	80	AA512632	AA512632 xx77c09.x
C 22	332	17.4	414	39	AI434224	AI434224 t133a06.x
C 23	329.2	17.2	367	33	AA991282	AA991282 os51a12.s
C 24	328.6	17.2	368	40	AI475832	AI475832 tc87a11.x
C 25	327.4	17.1	417	31	AA814368	AA814368 nz08a02.s
C 26	325.4	17.0	430	91	W86422	W86422 zh61a12.s1
C 27	321.8	16.8	419	44	AI813566	AI813566 wj83f08.x
C 28	318.6	16.7	365	36	AI224090	AI224090 qx13d07.x
C 29	318	16.6	398	39	AI439553	AI439553 tc90c06.x
C 30	316.2	16.6	355	39	AI433629	AI433629 t133f04.x
C 31	314.2	16.5	352	81	AW571597	AW571597 xx33g10.x
C 32	311.2	16.3	414	32	AA883421	AA883421 am25a04.s
C 33	311.2	16.3	430	31	AA832318	AA832318 oc95h05.s
C 34	305	16.0	387	39	AI439627	AI439627 tc91c06.x
C 35	303.4	15.9	312	79	AW402953	AW402953 UI-HF-BK0
C 36	297.8	15.6	340	38	AI383514	AI383514 tg03d02.x
C 37	295	15.4	340	78	AW337573	AW337573 xx59h01.x
C 38	287.2	15.0	518	82	AW655210	AW655210 105863 MA
C 39	273.2	14.3	341	35	AI083685	AI083685 oy71g07.x
C 40	266.2	13.9	322	63	AI865238	AI865238 wk08h10.x
C 41	265.2	13.9	382	34	AI057449	AI057449 ow80c02.x
C 42	259	13.6	494	82	AW655218	AW655218 105873 MA
C 43	257.2	13.5	369	31	AA832322	AA832322 oc95h11.s
C 44	254	13.3	287	35	AI083684	AI083684 oy71g06.x
C 45	254	13.3	482	79	AW444691	AW444691 UI-H-B13-

ALIGNMENTS

RESULT 1  
AI760373/c 1760373 711 bp mRNA EST 20-DEC-1999  
LOCUS wh87d11.x1 NCI-CGAP-CLL1 Homo sapiens cDNA clone IMAGE:2387733 3'  
DEFINITION similar to TR:060667 O60667 ANTI-FAS-INDUCED APOPTOSIS. ;, mRNA

sequence.  
AI760373  
AI760373.1 GI:5176040  
EST.  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 711)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On Oct 30, 1998 this sequence version replaced gi:3812483.  
Contact: Robert Strausberg, Ph.D.  
Tel.: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever,  
M.D., Louis M. Staudt, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.lnl.gov/bbrp/image/image.html](http://www-bio.lnl.gov/bbrp/image/image.html)  
Seq primer: -400P from Gibco  
Insert Length: 1449 Std Error: 0.00  
High quality sequence stop: 472.  
Location/Qualifiers  
1..711  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="IMAGE:2387733"  
/tissue\_type="B-cell, chronic lymphocytic leukemia"  
/lab\_host="DH10B"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer [5',  
TGTTACCACTGAGTGGGAGCGCGCATGCTTTTGTCTTTTGTCTTTT  
T 3']; double-stranded cDNA was ligated to Eco RI  
adaptors (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of the modified pT7T3 vector.  
Library is normalized, and was constructed by Bento  
Soares and M.Patima Bonaldo."  
BASE COUNT 127 a 195 c 239 g 148 t 2 others  
ORIGIN  
Query Match 33.3%; Score 635.8; DB 44; Length 711;  
Best Local Similarity 95.6%; Pred. No. 8.1e-164;  
Matches 674; Conservative 0; Mismatches 29; Indels 2; Gaps 2;  
Qy 641 ggtccagagcatcttcagtagcaggtgacagcccccgaacctctgcatccataca 700  
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 711 GTGTGAGAGCATTCAGTAGCAGGTGACAAAGCCGACCTCTTTGCCATCCTACA 652  
Qy 701 gctcaaaatctcagctctggaggggtgctcaagcccccagccagccagctaca 760  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 651 GCTCAAAAATCTCAGCTCTGGAGGGGTGCTCAAGCCCCAGACCCAGGCTTACACACA 592  
Qy 761 cacaccagggc-tgcacagggcagagcagctggactatggctcacagtctggaggaggaag 819  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 591 CACACAGCGCTGCACAGGCAGAGCAGTGGACTGTGCTCACAGTCTGGAGGGAAGG 532  
Qy 820 ccaagattcaactctatccgaccatctctggcccttttcctgctgggactctcg 879  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 531 CCAAGGATTTTTCATCTGATCCCGACCATCTCTGGGCTTTTCTGCTGGGCTCTGGG 472  
Qy 880 gctgggtgtaaaagggcgcttgaagagagaaagccctctccagggcgccgcgcgact 939  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 471 GCTGGTGTGAAAAGGGCCCTTCAAAGGAGGAAAGCCCTCTCCAGGCGGGCCGCCGACT 412











Db 349 TCTGGGATTGACAGAGTGTCTTCCATGTCTTTGACCACAGGGT-TGTTGCTGCCCAGG 291

QY 1588 ctctagatcacatgcatcagctcgtggcagagggatagctattgtctcgggcatccttc 1647

Db 290 CTCTAGATCACATGACATCAGCTGGGCGAGAGGCATAGCTATTGCTCGGGCATCCTTC 231

QY 1648 ccaggggtgggtcttacacaataaaggctcttctctgagttatgtgacgtgacctcag 1707

Db 230 CCAGGGTGGGTCTTACACAAATACAGAGCTCTTCTCTGAGTTATGTGACATGCCTCAG 171

QY 1708 ccccatgactaacgaggggtctggtat--aaacactcctggaacgaccttggccctgat 1765

Db 170 CCCCATGGAAGTAAAGAGGGGTCTGGTATATAAAACACACTCCTGGAAACGCTTTTGCCTGAT 111

QY 1766 ccaaatgttagcacttctagtgaaagcttacttctcctcaagtctctgctctaaaggcaat 1825

Db 110 CCAATGTAGCACTTGTAGTGAACGCTCTACTTATCTCAAGTTCTATGCTTAAAGGCAAT 51

QY 1826 ttatcttgatgtgatataaaccacacttattagcaagata 1866

Db 50 TTATCTTGATGTGATGATAAACCAAACTTATTAGCAAAAAA 10

RESULT 8

AA769441/c

LOCUS

DEFINITION

nz41f03.s1 NCI\_CGAP\_GCB1 Homo sapiens cDNA clone IMAGE:1290365 3', mRNA sequence.

ACCESSION

AA769441

VERSION

AA769441.1 GI:2820679

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 453)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

Unpublished (1997)

On Dec 18, 1997 this sequence version replaced gi:2339900.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman, Ph.D., Gerald Marti, M.D.

cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

Insert Length: 1366 Std Error: 0.00

Seq primer: -40m13 fwd. ET from Amersham

High quality sequence stop: 286.

Location/Qualifiers

1. .453

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:1290365"

/clone\_1ib="NCI\_CGAP\_GCB1"

/tissue\_type="germinal center B cell"

/lab\_host="DH10B"

/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker. Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (CD20+, IgD-), provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(dT) primer

[5'-TGTTACCAATCTGAAGTGGAGCGCGCCCTATTTTTTTTTTTT-3', ]. Double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 126 a 103 c 105 g 119 t

ORIGIN

Query Match 21.4%; Score 409.2; DB 30; Length 453;

Best Local Similarity 97.5%; Pred. No. 1.2e-101;

Matches 437; Conservative 0; Mismatches 8; Indels 3; Gaps 2;

QY 1439 qcagccccctgtagttgacaggtgaagctgtaggcatgtagagcaattgtcccaatgccac 1498

Db 453 GCAGCCCCCTGTAGTTGACAGGTAAGCTGTAGGCATGTAGAGCAATTGTCCCAATGCCAC 394

QY 1499 ttgctctcttccaaagccgtcgacagactgtggattgttcgacagtggtttcttccatgtc 1558

Db 393 TTGCTTCTTCCAAAGCGCTCGAACAGACTGTGGGATTTCACAGAGTGTTCCTCCATGTC 334

QY 1559 ttgaccacaggggtgtgtgtgtctgcaggctctagatccatcgatcaggctggggcag 1618

Db 333 TTTGACCACAGGGT-TGTTGCTGCCAGGGCTTAGATCACATGGCATCAGGCTGGGGCAG 275

QY 1619 aggcatagtctgtctcgggcatcctccagggttggtctctacacaaatagaaggtt 1678

Db 274 AGCATAGCTATTGCTCGGGCATCTCTCCAGGGTTGGGTCTTACACAAATAGAAAGGCT 215

QY 1679 cttgctctgagttatgtgacgtgcctcagccccatggactaaagcagggtccttggtat--a 1736

Db 214 CTTGCTCTGAGTTATGTGACATGCCTCAGCCCCATGCAGTACAGCGGGCTGGTATATAA 155

QY 1737 aacactctggaaaccccttggccctgataccaaattgagcactgtctagtgaacgtcta 1796

Db 154 AACACTCCTGGAAACGCCTTTGCCCTGATCCAAATGTAGCACTTGTAGTGAACGCTTA 95

QY 1797 cttatctcaagttctatgctaaaggcaattatcttgatgtatgataaaccacacttat 1856

Db 94 CTTATCTCAAGTTCTATGCTTAAAGGCANTTTATCTTGTGTGATGATGAACCAACTTAT 35

QY 1857 tagcaagatatgcataatatatccataaa 1884

Db 34 TAGCAAGATATGCATATATATCCATAA 7

RESULT 9

AW001912/c

LOCUS

DEFINITION

AW001912

IMAGE:2513908 3', mRNA sequence.

ACCESSION

AW001912

VERSION

AW001912.1 GI:5848828

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 441)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

Unpublished (1997)

On Oct 8, 1998 this sequence version replaced gi:3725720.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

This clone is available royalty-free through LLNL; contact the IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.

Seq primer: -40UP from Gibco

High quality sequence stop: 429.

Location/Qualifiers

1. .441

/organism="Homo sapiens"

COMMENT		On May 18, 1998 this sequence version replaced gi:3137192. Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert.Strausberg@nih.gov Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever, M.D., Louis M. Staudt, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/bbrp/image/image.html Insert Length: 2024 Std Error: 0.00 Seq primer: -40UP from Gibco High quality sequence stop: 453.	
FEATURES		Location/Qualifiers	
source		1..458	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/clone="NCE:2389821"	
		/clone_lib="NCI_CGAP_CLL1"	
		/issue_type="B-cell, chronic lymphocytic leukemia"	
		/lab_host="DH10B"	
		/note="Vector: p7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGAGCGCGCATGCTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p7T3 vector. Library is normalized, and was constructed by Bento Soares and M. Fatima Bonaldo."	
BASE COUNT		126 a 105 c 108 g 118 t 1 others	
ORIGIN			
Query Match		20.9%; Score 400; DB 44; Length 458;	
Best Local Similarity		96.8%; Pred. No. 4e-99;	
Matches 429; Conservative		0; Mismatches 11; Indels 3; Gaps 2;	
QY	1417	ccaggtcgtcttcgacacgttgccagccccctgttagtgacaggtgaagctgtagcatgt	1476
Db	458	CCAGGTCGCTTGCACACCTTGGCAGCCCCCTGTAGTTGACAGTAAGCTGTAGGCATGT	399
QY	1477	agagcaattgtcccaatgcactgtctcttcccaagccgtcgacagactgtggatt	1536
Db	398	AGAGCAATTGTCCCAATGCCACTTGCCTCTTCCAGCCGTCGAACAGACTGTGGGATT	339
QY	1537	tgcagagtgtttctccatgtcttgaccacaggggtgtgtgtgccaggtcttagatc	1596
Db	338	TGCAGAGTGTCTTCATGTCTTTGACCACAGGGT-TGTTGTCGCCAGGCTCTAGATC	280
QY	1597	acatggcatcaggtcgggagaggtatgttctcgtgggcctccctcccaagggtg	1656
Db	279	ACATGGCATCAGGCTGGGCGANAGGCATAGCTATTGTCTCGGCGCATCCTCCAGGGTTG	220
QY	1657	ggtcttacacaaataaagcgtctgtctgatttatgaactgacctcagccccatgga	1716
Db	219	GGTCTTACAAATAAGAGGCTCTGTCTCTGAGTTATGTGACATGCTCAGCCCCATGGA	160
QY	1717	ctaagcagggtctggtat--aaacactccttggaacgccttgcctcctgcatcaaatgt	1774
Db	159	CTAAGCAGGGGTTTGGTATATAAAACACTCTCTGAAACGCCCTTTCGCTGATCCAAATGT	100
QY	1775	agcactgttagtgaaactctactatctcaagttctatgctaaagcaattatctga	1834
Db	99	AGCACTTGTCTAGTGAACGCTCTACTATCTCAAGTCTTATGCTAAGCAAAATATCTTGA	40
QY	1835	tgtgatgataaaccaactatt	1857
Db	39	TGTCATGATAAACCAACTTATT	17

```

RESULT 11
AA456309/c
LOCUS
DEFINITION
  aa13q04.s1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:813174
  3', mRNA sequence.
ACCESSION
  AA456309
VERSION
  AA456309.1 GI:2179519
KEYWORDS
  EST.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
  1 (bases 1 to 476)
  Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
  Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J.,
  Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,
  White, Y., Wylie, T., Waterston, R. and Wilson, R.
  WashU-Merck EST Project 1997
  Unpublished (1997)
  Contact: wilson rk
  Washington University School of Medicine
  4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
  Tel: 314 286 1800
  Fax: 314 286 1810
  Email: est@watson.wustl.edu
  This clone is available royalty-free through LLNL ; contact the
  IMAGE Consortium (info@image.llnl.gov) for further information.
  Seq primer: -41ml3 fwd. ET from Amersham
  High quality sequence stop: 459.
FEATURES
  source
    Location/Qualifiers
      1..476
        /organism="Homo sapiens"
        /db_xref="GDB:6044084"
        /db_xref="taxon:9606"
        /clone="IMAGE:813174"
        /clone_lib="Soares_NhHMPu_S1"
        /tissue_type="pooled human melanocyte, fetal heart, and
        pregnant uterus"
        /lab_host="DH10B"
        /note="Organ: mixed (see below); Vector: pT7T3D-Pac
        (Pharmacia) with a modified polylinker; Site.1: Not I;
        Site.2: Eco RI; Equal amounts of plasmid DNA from three
        normalized libraries (melanocyte 2NDHM, pregnant uterus
        NBHPU, and fetal heart NBH19W) were mixed, and ss circles
        were made in vitro. Following HAP purification, this DNA
        was used as tracer in a subtractive hybridization
        reaction. The driver was PCR-amplified cDNAs from pools of
        5,000 clones made from the same 3 libraries. The pools
        consisted of I.M.A.G.E. clones 260232-265223,
        340488-345479, and 484488-489479."
BASE COUNT
  140 a 98 c 121 g 117 t
ORIGIN
  Query Match 20.9%; Score 399; DB 26; Length 476;
  Best Local Similarity 97.1%; Pred. No. 7.7e-99;
  Matches 428; Conservative 0; Mismatches 10; Indels 3; Gaps 2;
  QY 1472 catagagcaattgtcccaatgccactgtcttccttccaaagcgtgcaacagctgtg 1531
  Db 476 CATGTAGAGCAATGTCCCAATGCCACTTGTCTTCATGCTTTCCAGCCGTCGAACACTGTG 417
  QY 1532 ggatttgcagagtgtttcttccatgtttgtgaccacaggggtgtgtgtgtgctgccagctct 1591
  Db 416 GGATTTGCAGAGTGTCTTCTCCATGCTTTGACCAACAGGGT-TGTTGCTGCCAGGCTCT 358
  QY 1592 agatcacatggcatcaggtcgggacagagcatagctattgtctcgggcacatcttcccccag 1651
  Db 357 AGATCACATGGCATCAGGCTGGGGCAGAGCATAGCTATTGTCTCGGCATCTCTTCCCCAG 298
  QY 1652 ggttgaggtcttacacaaatagaaggcttcttgcctagttatgtcagtcgctcagcccc 1711
  Db 297 GGTGGGTCTTACACAAATAGAGGCTCTTGCTCTGAGTTATGTGACATGCCTCAGCCCC 238

```

---

```

QY 1712 atggactaacgaggggtctggtat--aaacactcttggaacgccttgcctgatccaa 1769
  Db 237 ATGGACTTAACGAGGGGTCTGGTATAAAACACACTCTGTGAAACGCTTTGGCCTGATCCAA 178
  QY 1770 atgttagcacttgcgtgaacgtctactctatctcaagttctatgctaaaggaatttat 1829
  Db 177 ATGTAGCACTTGTCTAGTGAACGCTACTTATCTCAAGTTCTATGCTAAAGGCAATTAT 118
  QY 1830 cttgagtgtgatataaacacaaactatttagcaagatgcatatataatccataaaattctc 1889
  Db 117 CTTGATGTGATGATAAACCAAACTTATTAGCAAGATGATCATATATATCCATAAATCTTC 58
  QY 1890 tttaactctgtctccatccttt 1910
  Db 57 TTTACTCTGTCTCCATCACTT 37
  RESULT 12
  AA703531/c
  LOCUS
  DEFINITION
    zj13h01.s1 Soares_fetal_liver_spleen_INFLS_S1 Homo sapiens cDNA
    clone IMAGE:450193 3', mRNA sequence.
  ACCESSION
    AA703531
  VERSION
    AA703531.1 GI:2713449
  KEYWORDS
    EST.
  SOURCE
    human.
  ORGANISM
    Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
    1 (bases 1 to 474)
    Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
    Kriaman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,
    Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,
    Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.
    WashU-NCI human EST Project
    Unpublished (1997)
    On Jan 19, 1998 this sequence version replaced gi:2045722.
    Contact: wilson rk
    Washington University School of Medicine
    4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
    Tel: 314 286 1800
    Fax: 314 286 1810
    Email: est@watson.wustl.edu
    This clone is available royalty-free through LLNL ; contact the
    IMAGE Consortium (info@image.llnl.gov) for further information.
    Seq primer: -40ml3 fwd. ET from Amersham.
FEATURES
  source
    Location/Qualifiers
      1..474
        /organism="Homo sapiens"
        /db_xref="GDB:1354450"
        /db_xref="taxon:9606"
        /clone="IMAGE:450193"
        /clone_lib="Soares_fetal_liver_spleen_INFLS_S1"
        /sex="male"
        /dev_stage="20 week-post conception fetus"
        /lab_host="DH10B (ampicillin resistant)"
        /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
        with a modified polylinker; Site.1: Pac I; Site.2: Eco RI;
        This is a subtracted version of the original Soares fetal
        liver spleen INFLS library. 1st strand cDNA was primed
        with a Pac I - oligo(dT) primer [5'
        AACTGGAGAAATTAATTAAGATCTTTTTTTTTTTTTTTTTT 3'],
        double-stranded cDNA was ligated to Eco RI adaptors
        (Pharmacia), digested with Pac I and cloned into the Pac I
        and Eco RI sites of the modified pT7T3 vector. Library
        went through one round of normalization. Library
        constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT
  138 a 98 c 120 g 118 t
ORIGIN
  Query Match 20.5%; Score 390.8; DB 30; Length 474;

```

Best Local Similarity 96.6%; Pred. No. 1.4e-96;  
Matches 421; Conservative 0; Mismatches 12; Indels 3; Gaps 2;

QY 1477 agacaatgtcccaatgcaactgtctcttcccaagcgtcgaaacactgtggatt 1536  
|||||  
Db 474 AGACAATGTCCCAATGCCACTTGTCTTCCCAAGCGTGAACACACTGCGGATT 415  
|||||  
QY 1537 tqcagaggtttctccatgtctttgaccacaggggtgtgtgtcgtccaggctctagatc 1596  
|||||  
Db 414 TGCAGAGTGTCTTCCCATGCTTTGACCACAGGCT-NGTTGTGCGCCAGGCTCTAGATC 356  
|||||  
QY 1597 acatgcatcagcgtggggcagaggaatagctattgtctcgggcatcttccagggttg 1656  
|||||  
Db 355 ACATGGCATCAGCGTGGGGCAGAGGCATAGCTATTGTCTCGGGCATCTTCCAGGGTTG 296  
|||||  
QY 1657 ggtctacacaataagaaggctctgtcgtgagttatgtgacgtgctcagccccatgga 1716  
|||||  
Db 295 GGTCTTACAAATAGAGGCTCTGTCTGTGATTTATGTGACATGCTCAGCCCCATGGA 236  
|||||  
QY 1717 ctaagcaggggtcgtgat--aaacactcctggaaccccttgcctcatccaaatgtt 1774  
|||||  
Db 235 CTAAGCAGGGGTCTGGTATAAAAAACACTCTGGAACGCCCTTTGGCCCTCATCAATGTT 176  
|||||  
QY 1775 agcactgtcagtgaaagctctacttatctcaagttctatgctaaaggcaatttatcttga 1834  
|||||  
Db 175 AGCACTTGTCTAGTGAACGCTACTTATCTCAAGTTCTATGCTAAAGCAATTTATCTTGA 116  
|||||  
QY 1835 tctgatataaaacaaacttattagaagatatgcataatatocataaattctctttac 1894  
|||||  
Db 115 TGTGATGATAAACCAAACTTATTAGCAAGATATGATATATATTCATTAATATCTTTAC 56  
|||||  
QY 1895 tctgtctccatcttt 1910  
|||||  
Db 55 TCTGTCTCCATCACTT 40

RESULT 13  
AW401870 410 bp mRNA EST 16-FEB-2000  
LOCUS  
DEFINITION  
UI-HF-BK0-aah-g-11-0-UI.r1 NIH\_MGC\_36 Homo sapiens cDNA clone  
IMAGE:3053781 5', mRNA sequence.  
ACCESSION  
AW401870  
VERSION  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 410)  
AUTHORS  
NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.  
TITLE  
National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL  
Unpublished (1999)  
COMMENT  
On Jun 22, 1998 this sequence version replaced gi:3246884.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: [Robert.Strausberg@nih.gov](mailto:Robert.Strausberg@nih.gov)  
Eco RI site shown at the beginning of the sequence.  
Tissue procurement: Louis M. Staudt, M.D., Ph.D.  
cDNA Library Preparation: M.B. Soares Lab  
cDNA Library Arrayed by: M.B. Soares Lab  
DNA Sequencing by: M.B. Soares Lab  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
Seq primer: M13 Forward.  
Location/Qualifiers  
1. .410  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3053781"  
/clone\_lib="NIH\_MGC\_36"  
/tissue\_type="lymph"  
/cell\_type="germinal center B cells"

/cell\_line="MGC85"  
/lab\_host="DH10B (LTI)"  
/note="Vector: p77T3-Pac; Site\_1: NotI; Site\_2: Eco RI;  
Constructed from size fractionated cytoplasmic mRNA  
(0.5-1.5kb). Directionally cloned. Cells provided by Louis  
M. Staudt, Ph.D. Library preparation by Maria de Fatima  
Bonaldo, Ph.D. and M. Bento Soares, Ph.D."

BASE COUNT 72 a 154 c 94 g 89 t 1 others  
ORIGIN

Query Match 20.0%; Score 382.4; DB 79; Length 410;  
Best Local Similarity 99.2%; Pred. No. 2.7e-94;  
Matches 394; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1041 tgaagcggacgtcgcagggcaggggagggcccccgttcccgcccgagcgagcggttgc 1100  
|||||  
Db 15 TGGAGCGGAGCGTGCAGGCACAGGGAGGCCCCCGTTCCCGCGCCGGAGCGCGTTGCC 74  
|||||  
QY 1101 cccgcgccgctgcaggtgtctgaattccctggctccatgcccctctctgaagaccag 1160  
|||||  
Db 75 CCCC GCCCGCTGCAGGTGTCTGAATCTCCCTGGCTCCATGCCCATCTCTGAAGACCAG 134  
|||||  
QY 1161 ctgtgaatacgtgagcgtctaccaccagcctgcccgcctatggaagacagtgattcaga 1220  
|||||  
Db 135 CTGTGAATACGTGAGCCTCTACCACCAGCCTGCCGCCCATGATGGAGGACGTGATTGAGA 194  
|||||  
QY 1221 tgactacataatgttctcgtcgtgacactcccaagctatcccccaacccagcgtcggg 1280  
|||||  
Db 195 TGACTACATCAATGTTCTGCCTGACAACTCCCGAGCTATCCCCAACCCAGGCTCGGA 254  
|||||  
QY 1281 ctgtgtgccaaggagtgctcatctatctgtcgtgatgccaatacctgtcttcattgtctc 1340  
|||||  
Db 255 CTGTGTGGCAAGGAGTGTCTATCTATCTNCTGATGTCCAAATGCTGCTTCATGTGTCTC 314  
|||||  
QY 1341 agagccctcatcacttcccatgcccctatcgcactcccccctctctatctgtggtccct 1400  
|||||  
Db 315 AGAGCCCTCATCACTTCCCATGCCCATCTGAGCTCCCATCCCATCTATCTGT-GCCT 373  
|||||  
QY 1401 gagcatggtctgtgccccaggtcgtcttgcacacctt 1437  
|||||  
Db 374 GAGCATGGCTCTTCCCCAGGTGCTGTGCACACCTT 410  
|||||

RESULT 14  
AA456707  
LOCUS  
DEFINITION  
aal3g04.r1 Soares\_NhMpu\_S1 Homo sapiens cDNA clone IMAGE:813174  
5', mRNA sequence.  
ACCESSION  
AA456707  
VERSION  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 404)  
AUTHORS  
Hallier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,  
Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J.,  
Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,  
White, Y., Wyllie, T., Waterston, R. and Wilson, R.  
WashU-Merck EST Project 1997  
Unpublished (1997)  
COMMENT  
On Jan 14, 1998 this sequence version replaced gi:1877745.  
Contact: Wilson R.  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: [est@watson.wustl.edu](mailto:est@watson.wustl.edu)  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Seq primer: -28m13 rev2 ET from Amersham



Align seg 1/1 to: X28179 from: 1 to: 1047

```

1 MetAspArgTIpLeuTTPProLeuTyrPheLeuProValSerGlyAlaLe 17
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
150 ATGGACTTCTGCTTTGGCCACTTTACTTCTGCGCATCAAGAGGGCCCT 199
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
17 uArglleLeuProGluValysValGluGlyLeuGlyGlySerValT 34
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
200 GAGGATCTCCAGAGTAAAGGTAGAGGGGGAGCTGGCGGATCAGTTA 249
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
34 hrileLysCysProLeuProGluMetHisValArglleTyrLeuCysArg 50
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
250 CCATCAAGTGGCCACTTCTGGAATGCAATGTGAGGATATATCTGTGCGG 299
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
51 GluMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsnPh 67
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
300 GAGATGGCTGGATCGGAACATGTGTACCGTGGTATCCACCACTT 349
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
67 eileLysAlaGlyTyrLysGlyArgValThrLeuLysGlnTyrProArgL 84
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
350 CATCAAGGCAAGTAAAGGGCGGAGTTTACTCTGAAGCAATACCAAGCA 399
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
84 ysAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerGly 100
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
400 AGAATCTCTTCCTAGTGGAGGTAAACACAGCTGACAGAAAGTACAGCGGA 449
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
101 ValTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLy 117
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
450 GTCTATGCTTGGCGAGC.GGCATGAACACAGACCGGGGAAAGACCCAGAA 498
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
117 sValThrLeuAsnValHisSerGluTyrGluProSerTyrGluGlnP 134
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
499 AGTCACCTCAATGTCCACAGTGAATACGAGCCATCATGGGAAGAGCAGC 548
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
134 toMetProGluThrProLysTyrPheHisLeuProTyrLeuPheGlnMet 150
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
549 CAATGCTGAGATCCAAATGGTTTCATCTGCCCTATTGTTCAGATG 598
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
151 ProAlaTyrAlaSerSerSerLysPheValThr..ArgValThrThrPro 166
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
599 CTTGATATGCCGGTCTTCCACATTCGTAAACCGAGAGTTACCAACCA 648
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
167 Ala.GlnArgGlyLysValProValHisHisSerSerProThrThrG 183
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
49 GCTTAAAGGGGCAAGGTCCTCCAGTTCACCACTCTCCGCCACCAACC 698
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
183 Inile.ThrHisArgPro.ArgValSerArgAlaSerSerValAlaGlyA 199
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
699 AAATTCACCAACCGCCCTTCGAGTNCAGAGCATCTTCAGTAGCAGGTG 748
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
199 spLysProArg..ThrPheLeuProSerThrThrAlaSerLysLysSerAl 215
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
749 ACAAGCCCCGAAACTTCTGCCATCCACTACAGCCCTCAAAATCTCAGC 798
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
215 aLeuGluGlyLeuLeuLys...ProGlnThrProSerTyrAsnHisHist 231
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
799 TCTGGAAGGCTGCTTCAAGCCCCAGAGGCCCGCCAGCTACACAN..CACA 847
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
231 hrArgLeuHisArgGlnArgAlaLeuAsp...Tyr.GlySerGlnSerG 246
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
848 CCAGCTTGCACAGGAGAGAGCACTGGATACTTATGGGNTCAGAGTCTGG 897
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
246 yArgGluGly.....GlnGlyPhe...HisIleLeuIlePro..Thril 259
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
898 G...GAGGGAANGNCCAAAGGATTTTNAACATTCCTGATTCGCGGACCATC 944
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
260 LeuLysLeuPhe 263
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
945 NTGGGGGCTTT 956
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

seq\_name: N\_Geneseq\_36:T31290

```

seq_documentation_block:
ID T31290 standard; cDNA; 3095 BP.
AC T31290;
DE Mouse poly-immunoglobulin receptor, cDNA.
KW Mouse; immunoglobulin; receptor; protection protein; mutants;
KW heavy chain; antigen binding domain; protection; pathogen;
KW mucosal; environment; gastrointestinal; passive; immunisation;
KW Guy's 13 antibody; prevention; dental caries; Streptococcus;
KW poly; sorbinus; murine; ss.
OS Mus musculus.
FH Key Location/Qualifiers
FT cds 85..2400 /*tag= a
FT W09621012-A1.
PD 11-JUL-1996.
PF 27-DEC-1995; U16889.
PR 30-DEC-1994; US-367395.
PR 04-MAY-1995; US-434000.
PA (PLAN-) PLANT BIOTECHNOLOGY INC.
PA (UNME-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.
PA (PLAN-) PLANET BIOTECHNOLOGY INC.
PI Hatt AC, Lehner T, Ma JKC;
DR WPI; 96-333987/33.
DR P-PSDB; W03180.
PT Immunoglobulin and protection protein complex and its prodn. in
PT plants - useful for passive immunisation against mucosal antigens,
PT esp. against S. mutans and S. sorbinus to prevent dental caries
PS Disclosure: Pages 117-121: 152pp; English.
CC The present sequence encodes the mouse poly-immunoglobulin (Ig)
CC receptor, a portion of which corresp. to residues 1-627, pref.
CC 1-608, or esp. residues 13-45, 1-120, 110-230, 210-340, 320-450,
CC 440-550, 550-606 or 550-627 comprises a protection protein (PP).
CC The Ig of the invention comprises a PP as above in association with
CC an Ig derived heavy chain, having at least a portion of an antigen
CC (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.
CC gastrointestinal, environments, therefore enhancing its
CC effectiveness in passively immunising animals against mucosal
CC pathogens. The Ag binding domain is specifically derived from the
CC Guy's 13 antibody, and the Ig can be used to prevent dental caries
CC by binding, e.g. Streptococcus mutans serotypes c, e and f, or
CC S. sorbinus serotypes d and g.
SQ Sequence 3095 BP; 861 A; 796 C; 784 G; 654 T;

```

alignment\_scores:  
Quality: 198.00 Length: 168  
Ratio: 1.980 Gaps: 5  
Percent similarity: 59.524 Percent Identity: 31.548

alignment\_block:  
US-09-135-238B-2 x T31290 ..

Align seg 1/1 to: T31290 from: 1 to: 3095

```

10 PheLeuProValSerGlyAlaLeuArglleLeuProGluValLysValG 26
||| ||||| ||||| |||||
121 TTTTACGGGGTCTCCACAAAAGCCCATATTGTGTCCTCCAGGAGTGAG 170
||||| ||||| ||||| |||||
26 uGlyGluLeuGlyGlySerValThrLysCysProLeuProGlu.... 41
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
171 TAGTATAGAGCGGAGCTCTGTTCCATCATCGTGTACTACCCAGACCT 220
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
42 .....MetHisValArglleTyrLeuCysArgGluMetAlaGlySer 55
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
221 CTGTCACCGGCACACCCGGAATATCTGTGTCGCCAAGGAGCC...AGC 267
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
56 GlyThrCysGlyThrValValSerThrThrAsnPheIleLysAlaGlu 72
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
268 GGCATGTGCACACGCTCATCTCTCAATGGCTACCTCTCCAGGAGTA 317
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
72 rLysGlyArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeu 89
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```



334 roValProGlyProGlyAlaProLeuProProAlaProLeuGlnValSer 350  
|||||  
1146 CGGTTCGGCGCGGAGCGCGTTGGCCCGCGCGCTGCAGGTGCT 1195  
351 GluSerProTrpLeuHisAlaProSerLeuLysThrSerCysGluTrpVa 367  
|||||  
1196 GAATCTCCCTGGCTCCATGCCCATCTCTGAAGACCGAGCTGGAATACGT 1245  
367 lSerLeuTrpHisGlnProAlaAlaMetMetGluAspSerAspSerAsp 384  
|||||  
1246 GAGCTCTACACACAGCGCTGCCGCCATCATGGAGGACAGTATTTCAGATG 1295  
384 spTyrlleAsnValProAla 390  
|||||  
1296 ACTACATCAATGTTCTTCGCC 1315

seq\_name: N\_Geneseq\_36:X28179

seq\_documentation\_block:

ID X28179 standard; CDNA; 1047 BP.

AC X28179;

DT 16-JUN-1999 (first entry)

DE Human PIGRL-1 coding sequence fragment.

KW PIGRL-1; human; autoimmune disease; hyper-IgM Immunodeficiency; HIM;

KW X-linked Severe Combined Immunodeficiency; XSCID; IgA deficiency;

KW diagnosis; therapy; ss.

OS Homo sapiens.

PN EP-905238-A2.

PD 31-MAR-1999.

PF 14-AUG-1998; 306487.

PR 30-OCT-1997; US-961564.

PR 25-AUG-1997; US-056935.

PI (SMK ) SMITHKLINE BEECHAM CORP.

PI Sweet RW, Truneh A, Wu S;

DR WPI: 99-192666/17.

DR P-PSDB; 195002.

PT New polypeptides encoding PIGRL-1 useful for treating diseases such

PT as X-linked Severe Combined Immunodeficiency

PS Disclosure: Page 8; 26pp; English.

CC This sequence encodes the human PIGRL-1 protein of the invention.

CC Autimmune diseases involving altered expression or activity of PIGRL-1

CC may include Hyper-IgM Immunodeficiency (HIM), X-linked Severe Combined

CC Immunodeficiency (XSCID) and IgA deficiency. These diseases can be

CC diagnosed or susceptibility to them predicted by: (1) determining whether

CC there is a mutation in the genomic copy of the gene encoding PIGRL-1; or

CC (2) measuring the amount of PIGRL-1 in a sample derived from the patient.

CC Patients deficient in PIGRL-1 can be treated by administering either the

CC PIGRL-1 DNA or its complement or an agonist of PIGRL-1 to the patient.

CC Patients with excessive expression or activity of PIGRL-1 can be treated

CC by administering an antagonist of PIGRL-1, an antisense nucleic acid

CC molecule which inhibits the expression of PIGRL-1 or administering

CC sufficient PIGRL-1 to compete with the endogenous activity. PIGRL-1 can

CC be used to identify its agonists by contacting a cell expressing PIGRL-1

CC with a candidate compound in the presence of a signal system and noting

CC the candidate as an agonist if a signal is produced. The same method can

CC be used to identify antagonists of PIGRL-1 but the presence of an

CC antagonist is indicated by a decrease in production of the signal.

CC Antibodies against PIGRL-1 may be used to isolate or identify clones

CC expressing PIGRL-1. Polynucleotides encoding PIGRL-1 may be used to

CC identify chromosomal mutations in the gene encoding PIGRL-1 in patients.

CC This information may then be correlated with the incidence of autoimmune

CC disease in those patients to identify whether the mutation causes the

CC disease.

SQ Sequence 1047 BP; 255 A; 281 C; 276 G; 223 T;

alignment\_scores:

Quality: 1125.00 Length: 275

Ratio: 4.429 Gaps: 12

Percent Similarity: 92.364 Percent Identity: 90.182

alignment\_block:

US-09-135-238B-2 x X28179

34 hrIleLysCysProLeuProGluMetHisValArgIleLeuLysCysArg 50  
|||||  
246 CCATCAAGTGGCCCATCTCTGAAATGTCATGTGAGGATATATCTGTGCGCG 295  
51 GluMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsnPh 67  
|||||  
296 GAGATGGCTGGATCTGGAACATGTGTACCTGGTGTATCCACCAACCACTT 345  
67 eLleLysAlaGlyThrLysGlyArgValThrLeuLysGlnThrProArgL 84  
|||||  
346 CATCAAGGCGAATAACAAGGCGGAGTTACTCTGAAGCAATACCCACGCA 395  
84 ysAsnLeuPheLeuValLysValThrGlnLeuThrGluSerAspSergly 100  
|||||  
396 AGAATCTGTTCTAGTGGAGGTAAACACAGCTGACAGAAAGTGCAGCGGA 445  
101 ValTyAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnly 117  
|||||  
446 GTCTATGCTCGGAGGCGCATGAACACAGCGCGGGAAGACCCAGAA 495  
117 sValThrLeuAsnValHisSerGluThrGluProSerTrpGluGluGlnP 134  
|||||  
495 AGTCACCTGGAATGTCACAGTCAATACGAGCCATCATGGGAAGAGCAGC 545  
134 roMetProGluThrProLysTrpPheHisLeuProTyTrpLeuPheGlnMet 150  
|||||  
546 CAATGCTGAGACTCCAAATGTTTTCATCTGCCCTATTGTTCCAGATG 595  
151 ProAlaTyAlaSerSerSerLysPheValThrArgValThrThrProAl 167  
|||||  
596 CCTGCATATGCCAGTTCTTCCAAATTCGTAACCAAGAGTTTACCACACAGC 645  
167 aGlnArgGlyLysValProValHisHisSerSerProThrThrGlnI 184  
|||||  
646 TCAAGGCGCAAGGTCCTCAGTTTCCAGAGCATCTTCAGTAGCAGGTGACAAG 695  
184 leThrHisArgProArgValSerArgAlaSerSerValAlaGlyAspLys 200  
|||||  
696 TCACCCAGCGCCCTCGAGTGTCCAGAGCATCTTCAGTAGCAGGTGACAAG 745  
201 ProArgThrPheLeuProSerThrThrAlaSerLysIleSerAlaLeuGl 217  
|||||  
746 CCCCAGACCTTCTCCATCCTACCTACAGCTTCAAAATCTCAGCTCTGGA 795  
217 uGlyLeuLeuLysProGlnThrProSerTyTrpAsnHisHisThrArgLeuH 234  
|||||  
796 GGGGCTGCTCAAGCGCCAGAGCGCCAGCTACAAACACACACACAGGCTGC 845  
234 isArgGlnArgAlaLeuAspTyTrpGlySerGlnSerGlyArgGluGln 250  
|||||  
845 ACAGGAGAGAGACATGGACTATGGCTCACAGTCTGGGAGGGAGGCCAA 895  
251 GlyPheHisIleLeuIleProThrIleLeuGlyLeuPheLeuAlaLe 267  
|||||  
896 GGATTTTCACATCTGATCCCGACCATCTTGGGCCCTTTTCTGCTGGCACT 945  
267 uLeuGlyLeuValValLysArgAlaValGluArgGlyAlaLeuSerA 284  
|||||  
946 TCTGGGCTGGTGGTGAAGGCGCTTGAAGGAGGAAAGCCCTCTCCA 995  
284 rgArgAlaArgArgLeuAlaValArgMetArgAlaLeuGluSerSergln 300  
|||||  
996 GGGGCGCCCGCGAGCTGGCGGTGAGGATCGCGCCCTGGAGAGCTCCAG 1045  
301 ArgProArgGlySerProArgProArgSerGlnAsnIleTySerAl 317  
|||||  
1046 AGGCGCGCGGGTGCCTGCGGAGCGGCTCCCAAAACACATCTACAGCC 1095  
317 aCysProArgAlaArgGlyAlaAspAlaGlyThrGlyGluAlaP 334  
|||||  
1096 CTGCGCGCGCGCTCTGGAGCGGAGCGCTGAGGACAGGAGGCGCC 1145





OM of: US-09-135-238B-2 to: N\_Geneseq\_36:\* out\_format : pfs

Date: Sep 12, 2000 6:55 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL=frank_p2n.model -DEV=rlp  
-Q=/cygn2.1/USPTO.spool/US09135233/runat_24082000_091324_17918/app_query.fasta_1.1282  
-DB=N_Geneseq_36 -FMT=fastap -SUFFIX=p2n.rng -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-GAPOP=6.000 -GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct  
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM-ext -MNTLEN=0 -MAXLEN=100000  
-USER=US09135238 -CGN1_1_133 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPYX  
-WAIT -THREADS=1
```

Search information block:

```
Query: US-09-135-238B-2  
Query length: 390  
Database: N_Geneseq_36:*  
Database sequences: 311585  
Database length: 125096042  
Search time (sec): 121.040000
```

score\_list:

Sequence	Strd Orig	zScore	Escore	Len	Documentation
N_Geneseq_36:X28178	+	2047.00	1129.34	5.3e-111	2040
N_Geneseq_36:X28179	+	1125.00	1174.09	8.5e-58	1047
N_Geneseq_36:X28180	+	198.00	199.21	0.0017	3095
N_Geneseq_36:X28181	+	188.00	170.27	0.0697	29392
N_Geneseq_36:X28182	+	187.50	192.24	0.0042	1911
N_Geneseq_36:X28183	+	185.50	185.74	0.0096	3269
N_Geneseq_36:X28184	+	185.00	189.53	0.0059	1936
N_Geneseq_36:X28185	+	182.00	186.82	0.0083	1839
N_Geneseq_36:X28186	+	182.00	186.01	0.0093	2031
N_Geneseq_36:X28187	+	182.00	183.02	0.0136	2919
N_Geneseq_36:X28188	+	182.00	181.23	0.0171	3630
N_Geneseq_36:X28189	+	171.00	174.33	0.0414	2084
N_Geneseq_36:X28190	+	171.00	172.62	0.0515	2566
N_Geneseq_36:X28191	+	166.50	170.51	0.0676	1876
N_Geneseq_36:X28192	+	166.50	165.34	0.1312	3517
N_Geneseq_36:X28193	+	159.50	167.34	0.1015	1137
N_Geneseq_36:X28194	+	145.00	148.11	1.20	1876
N_Geneseq_36:X28195	+	143.50	133.93	7.37	8690
N_Geneseq_36:X28196	+	142.00	150.90	0.8361	914
N_Geneseq_36:X28197	+	139.50	130.12	12.02	8321
N_Geneseq_36:X28198	+	138.50	146.63	1.44	985
N_Geneseq_36:X28199	+	138.50	146.63	1.44	985
N_Geneseq_36:X28200	+	136.50	149.12	1.05	565
N_Geneseq_36:X28201	+	136.50	140.14	3.32	1683
N_Geneseq_36:X28202	+	136.50	123.79	27.06	11529
N_Geneseq_36:X28203	+	135.00	123.35	28.60	10704
N_Geneseq_36:X28204	+	131.00	118.16	55.71	12132
N_Geneseq_36:X28205	+	130.50	133.70	7.59	1617
N_Geneseq_36:X28206	+	129.50	134.18	7.13	1431
N_Geneseq_36:X28207	+	127.00	128.05	15.66	2197
N_Geneseq_36:X28208	+	126.50	120.88	39.26	4926
N_Geneseq_36:X28209	+	126.50	114.03	94.63	11336
N_Geneseq_36:X28210	+	125.50	129.22	13.48	1576
N_Geneseq_36:X28211	+	125.50	122.97	30.35	3400
N_Geneseq_36:X28212	+	125.50	116.64	67.64	7266
N_Geneseq_36:X28213	+	124.50	106.45	69.37	6357
N_Geneseq_36:X28214	+	124.50	118.70	187.39	16812
N_Geneseq_36:X28215	+	124.00	125.90	20.63	1951
N_Geneseq_36:X28216	+	124.00	124.32	25.27	2364

7073 ! TF8-5G9 CDR-grafted heavy c  
1929 ! Sorghum bicolor (L.) Moench  
1795 ! Human kidney injury related  
8051 ! Sugar biosynthesis gene clu  
8160 ! S-erythraea erythromycin-sy

seq\_name: N\_Geneseq\_36:X28178

seq\_documentation\_block:

ID X28178 standard; CDNA; 2040 BP.  
AC X28178;  
DT 16-JUN-1999 (first entry)  
DE Human PIGRL-1 coding sequence.  
KW PIGRL-1; human; autoimmune disease; hyper-igm Immunodeficiency; HIM;  
KW X-linked Severe Combined Immunodeficiency; XSCID; IGA deficiency;  
KW diagnosis; therapy; ss.  
OS Homo sapiens.  
PN EP-905238-A2.  
PD 31-MAR-1999. 306487.  
FF 14-AUG-1998; US-961564.  
PR 30-OCT-1997; US-056935.  
PR 25-AUG-1997; US-056935.  
PA (SMIK ) SMITHKLINE BEECHAM CORP.  
PI Sweet RW, Truneh A, Wu S;  
WP1: 99-192666/17.  
P-PSDB: Y05001.  
DR New polypeptides encoding PIGRL-1 useful for treating diseases such  
as X-linked Severe Combined Immunodeficiency  
PT Claim 4; Page 18-19; 26pp; English.  
PS This sequence encodes the human PIGRL-1 protein of the invention.  
CC Autoimmune diseases involving altered expression or activity of PIGRL-1  
CC may include Hyper-igm Immunodeficiency (HIM), X-linked Severe Combined  
CC Immunodeficiency (XSCID) and IGA deficiency. These diseases can be  
CC diagnosed or susceptibility to them predicted by: (1) determining whether  
CC there is a mutation in the genomic copy of the gene encoding PIGRL-1; or  
CC (2) measuring the amount of PIGRL-1 in a sample derived from the patient.  
CC Patients deficient in PIGRL-1 can be treated by administering either the  
CC PIGRL-1 DNA or its complement or an agonist of PIGRL-1 to the patient.  
CC Patients with excessive expression or activity of PIGRL-1 can be treated  
CC by administering an antagonist of PIGRL-1, an antisense nucleic acid  
CC molecule which inhibits the expression of PIGRL-1 or administering  
CC sufficient PIGRL-1 to compete with the endogenous activity. PIGRL-1 can  
CC be used to identify its agonists by contacting a cell expressing PIGRL-1  
CC with a candidate compound in the presence of a signal system and noting  
CC the candidate as an agonist if a signal is produced. The same method can  
CC be used to identify antagonists of PIGRL-1 but the presence of an  
CC antagonist is indicated by a decrease in production of the signal.  
CC Antibodies against PIGRL-1 may be used to isolate or identify clones  
CC expressing PIGRL-1. Polynucleotides encoding PIGRL-1 may be used to  
CC identify chromosomal mutations in the gene encoding PIGRL-1 in patients.  
CC This information may then be correlated with the incidence of autoimmune  
CC disease in those patients to identify whether the mutation causes the  
CC disease.

Sequence 2040 BP: 478 A: 616 C: 494 G: 452 T;

alignment\_scores:

Quality: 2047.00 Length: 390  
Ratio: 5.252 Gaps: 0  
Percent Similarity: 99.744 Percent Identity: 99.744

alignment\_block:

US-09-135-238B-2 x X28178

Align seg 1/1 to: X28178 from: 1 to: 2040

1 MetAspArgTrpLeuTrpProLeuTyrPheLeuProValSerGlyAlaLe 17

||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

146 ATGACCTCTGGCTTTGGCCACTTACTTCTCCAGTATCAGGGCCCT 195

||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

17 uArgIleLeuProGluValIysValGluGluLeuGlyGlySerValt 34

||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

196 GAGGATCTCTCCAGAGTAAGTAGAGGGGGGAGCTGGCGGATCAGTTA 245







```

* 31620 32323: contig of 704 bp in length
* 32324 32423: gap of 100 bp
* 32424 33110: contig of 687 bp in length
* 33111 33210: gap of 100 bp
* 33211 33902: contig of 692 bp in length
* 33903 34002: gap of 100 bp
* 34003 34708: contig of 706 bp in length
* 34709 34808: gap of 100 bp
* 34809 35500: contig of 692 bp in length
* 35501 35600: gap of 100 bp
* 35601 36300: contig of 700 bp in length
* 36301 36400: gap of 100 bp
* 36401 37097: contig of 697 bp in length
* 37098 37197: gap of 100 bp
* 37198 37900: contig of 703 bp in length
* 37901 38000: gap of 100 bp
* 38001 38695: contig of 695 bp in length
* 38696 38795: gap of 100 bp
* 38796 39487: contig of 692 bp in length
* 39488 39587: gap of 100 bp
* 39588 40291: contig of 704 bp in length
* 40292 40391: gap of 100 bp
* 40392 41101: contig of 710 bp in length
* 41102 41201: gap of 100 bp
* 41202 41886: contig of 685 bp in length
* 41887 41986: gap of 100 bp
* 41987 42686: contig of 700 bp in length
* 42687 42786: gap of 100 bp
* 42787 43484: contig of 698 bp in length
* 43485 43584: gap of 100 bp
* 43585 44276: contig of 692 bp in length
* 44277 44376: gap of 100 bp
* 44377 45078: contig of 702 bp in length
* 45079 45178: gap of 100 bp
* 45179 45869: contig of 691 bp in length
* 45870 45969: gap of 100 bp
* 45970 46664: contig of 695 bp in length
* 46665 46764: gap of 100 bp
* 46765 47456: contig of 692 bp in length
* 47457 47556: gap of 100 bp
* 47557 48249: contig of 693 bp in length
* 48250 48349: gap of 100 bp
* 48350 49042: contig of 693 bp in length
* 49043 49142: gap of 100 bp
* 49143 49843: contig of 701 bp in length
* 49844 49943: gap of 100 bp
* 49944 50640: contig of 697 bp in length
* 50641 50740: gap of 100 bp
* 50741 51438: contig of 698 bp in length
* 51439 51538: gap of 100 bp
* 51539 52231: contig of 693 bp in length
* 52232 52331: gap of 100 bp
* 52332 53020: contig of 689 bp in length
* 53021 53120: gap of 100 bp
* 53121 53857: contig of 737 bp in length
* 53858 53957: gap of 100 bp
* 53958 54652: contig of 695 bp in length
* 54653 54752: gap of 100 bp

```

## alignment\_scores:

```

Quality: 284.00      Length: 80
Ratio: 4.733        Gaps: 2
Percent Similarity: 75.000  Percent Identity: 73.750

```

## alignment\_block:

```

US-09-135-238B-2 x AC027719/rev

```

```

Align seg 1/1 to reverse of: AC027719 from: 1 to: 58628

```

```

2 AspArgTrpLeuTrp..... 6
||| ||| |||
40676 GACGCTGGAGCTGGAGTAGATGGCTGGATAGTATAGTAGATGGG 40627

```

```

7 .....ProLeuTyrPheLeuPro.....ValSerGlyAlaLeuArgI 19
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
40626 GTCCTCTGCATTAATCTCTCTCCCTGCAGTATCAGGGCCCTGAGGA 40577

19 leLeuProGluValValGluGluGluLeuGlySerValThrIle 35
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
40576 TCCTCCCAAGAAAGTAGAGGGGAGCTGGCGGATCAGTTACCATC 40527

36 LysCysProLeuProGluMethHisValArgIleTyrLeuCysArgGluMe 52
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
40526 AAGTGGCCACTTCTCTGAAATGATGTGAGGATATATCTGTGCCGGAGAT 40477

52 talaGlySerGlyThrCysGlyThrValValSerThrThr 65
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
40476 GGCTGGATCTGGAACATGTGTACCGTGTATCCACCACC 40437

seq_name: gb_ro:MMU06431

seq_documentation_block:
LOCUS MMU06431 3095 bp mRNA ROD 26-MAY-1995
DEFINITION Mus musculus B6/CBAF1J polymeric immunoglobulin receptor mRNA,
complete cds.
ACCESSION U08431
VERSION U08431.1 GI:458421
KEYWORDS complete cds.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 3095)
Piskurich,J.F., Blanchard,M.H., Youngman,K.R., France,J.A. and
Kastel,C.S.
Molecular cloning of the mouse polymeric Ig receptor. Functional
regions of the molecule are conserved among five mammalian species
J. Immunol. 154 (4), 1735-1747 (1995)
2 (sites)
Piskurich,J.F., Hsieh,M., Youngman,K.R. and Kastel,C.S.
Molecular cloning of mouse polymeric Ig receptor
J. Immunol. 150, 38A-38A (1993)
3 (bases 1 to 3095)
Piskurich,J.F.
Molecular Cloning and Regulation of the Polymeric Immunoglobulin
Receptor
Thesis (1994) Pathology, Case Western Reserve University
4 (bases 1 to 3095)
Piskurich,J.F.
Direct Submission
Submitted (08-FEB-1994) Janet F. Piskurich, Case Western Reserve
University, Pathology, Biomedical Research Building, Cleveland, OH
44060, USA

```

## FEATURES

## source

```

1..3095 Location/Qualifiers
/oranism="Mus musculus"
/strain="B6/CBAF1J (C57 Black 6 X CBA)"
/db_xref="taxon:10090"
/clone="G20"
/clone_lib="Catalog #935302, Stratagene Cloning Systems,
La Jolla, CA"
/sex="female"
/tissue_type="liver"
<1..84
85..138
85..2400
/codon_start=1
/function="binds and transports polymeric immunoglobulin"
/product="polymeric immunoglobulin receptor"
/protein_id="AAA67440.1"
/db_xref="GI:458422"
/translation="MRLVLTLLVTFSGVSTKPIFGPOEVSSTIGDSVSTCYVPD
TSSVHTRYKRCQACSCMCTLLSSNGYLSRGRANLIPENNTFVNIQLTQ
DSTGSKGLCTSNRGLSFDSYQVDELPDSRHYTIDGRVNTIEGPKRENPV

```

```

5'UTR
sig_peptide
CDS

```

34 roArgProArgSerGlnAsnAsnIleTyrSerAlaCysProArgAla 50  
 1062 CGCAGCCGCGCTCCCAAAACATCTACAGCGCTGCGCGCGCT 1111  
 51 ArgGlyAlaAspAlaLaglyThrGlyGluAlaProValProGlyProG1 67  
 1112 CGTGGAGCGACGCTGCAGGCACAGGGAGGCCCGCTTCCGCGCCCG 1161  
 67 yAlaProLeuProProAlaProLeuGlnValSerGluSerProTriLeuH 84  
 1162 AGCGCGCTGCGCCCGCGCGCTGAGGTCTGAATCTCCCTGGCTCC 1211  
 84 iAlaProSerLeuLysThrSerCysGluTyrValSerLeuTyrHisGln 100  
 1212 ATGCCCATCTCTGAAGACAGCTGTGAATACGTGAGCGCTCTACACAG 1261  
 101 ProAlaAlaMetMetGluAspSerAspAspTyrIleAsnValPr 117  
 1262 CTGCGCGCATGTGAGGACAGTGTATTCAGATGACTACATCATGTTC 1311  
 117 oala 118  
 1312 TGCC 1315  
 seq\_name: N\_Geneseq\_36:T93610

seq\_documentation\_block:  
 ID T93610 standard; DNA; 3946 BP.

DE 27-APR-1998 (first entry)  
 KW Mycobacterium tuberculosis genomic DNA fragment (I).  
 KW Tuberculosis; mycobacteria; infection; diagnosis;  
 KW antimycobacterial; antibiotic; vaccine; ss.  
 OS Mycobacterium tuberculosis.  
 PN W09741252-A2.  
 PD 06-NOV-1997.  
 PF 18-APR-1997; E01973.  
 PR 29-APR-1996; DE-017184.  
 PA (GFBF ) GFB GES BIOTEC FORSCHUNG GMBH.  
 PI Espitia C, Honisch C, Moreno C, Singh M;  
 DR WPI; 97-549750/50.  
 DR P-PSDB: W31852-55.  
 PT New DNA and related proteins or RNA derived from M. tuberculosis -  
 PT used for diagnosis of mycobacterial infections, monitoring  
 PT vaccination and development of anti-mycobacterial agents  
 PS Claim 1; Fig 9; 55pp; English.  
 CC New claimed DNA (A) has one of 3 isolated Mycobacterium  
 CC tuberculosis DNA sequences of 3946 bp (I) (T93610), 2653 bp (VI)  
 CC (see T93611) or 440 bp (IX) (see T93612), or hybridises to, or is a  
 CC fragment of these sequences. Also claimed are RNA transcripts  
 CC of (A), and proteins encoded by (A). To isolate (I), a cosmid  
 CC library of M. tuberculosis DNA was screened with degenerate probes  
 CC (see T93621) containing GC-rich regions and designed to encode part  
 CC of a proline-rich protein. (I) contains open reading frames (ORFs)  
 CC for 4 proteins (see W31852-55). These ORFs can be amplified and  
 CC cloned into vectors for expression in E. coli or other hosts.  
 CC Clones (VI) and (IX) were obtained from a phage library and  
 CC together encode 3 proteins (see W31851 and W31856-57). The claimed  
 CC DNA sequences can be used for diagnosing tuberculosis and other  
 CC mycobacterial infections in humans or animals, for identifying  
 CC mycobacteria in (clinical) samples by hybridisation or  
 CC as for epidemiological studies, for differentiation between strains,  
 CC the development of anti-mycobacterial drugs and vaccines.  
 SQ Sequence 3946 BP; 518 A; 1541 C; 1345 G; 542 T;

alignment\_scores:  
 Quality: 101.50 Length: 107  
 Ratio: 1.750 Gaps: 6  
 Percent Similarity: 54.206 Percent Identity: 36.449

alignment\_block:  
 US-09-135-238b-2\_copy\_273\_390 x T93610 ..  
 Align seg 1/1 to: T93610 from: 1 to: 3946  
 1 LysArgAlaValGluArgArgLys.....Al 9  
 1617 GCGCGTTCGCGCGCGCGCGCGCGCTTCTGCTGATGCTGTAACT 1666  
 9 aLeuSerArgAlaArgAlaArgLeuAlaValArgMetArgAlaLeuGluS 26  
 1667 GCTGCCCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGAGAG 1716  
 26 erSerGlnArgProArgGlySerProArgProArgSerGlnAsnAlle 42  
 1717 TAGGCGCGCTTCCCGCGGTTCGCGCGCGCGCGCGCGCGCGCGAGGAGCTG 1766  
 43 TyrSerAlaCysProArgAlaArg.GlyAlaAspAla.....AlaG 57  
 1767 CGCCACCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGAGTAGGCG 1816  
 57 lyThrGlyGluAlaProVal.....ProGlyPro..... 66  
 1817 GCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGTGATCCGACCTCCCGA 1866  
 67 .....GlyAla.ProLeuProAlaProLeuGlnValSerGluSerP 81  
 1867 GCG 1916  
 81 roTrpLeuHisAlaPro 86  
 1917 GGTTCGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1933  
 seq\_name: N\_Geneseq\_36:T93610

seq\_documentation\_block:

ID T93610 standard; CDNA; 266 BP.  
 AC T93610; 266 BP.  
 DT 12-FEB-1999 (first entry)  
 DE EST clone EAL10.  
 KW Expressed sequence tag; secreted protein; haematopoiesis regulator;  
 KW tissue growth; activin; inhibitor; tumour invasion suppressor; EST; human;  
 KW chemotaxis; chemokines; haemostasis; gene therapy; thrombolysis;  
 KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.  
 OS Homo sapiens.  
 PN W09845437-A2.  
 PD 15-OCT-1998.  
 PF 10-APR-1998; U06956.  
 PR 10-APR-1997; US-837312.  
 PA (GEMY ) GENETICS INST INC.  
 PI Acostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,  
 PI Racie LA, Spaulding V, Treacy M;  
 DR WPI; 99-070078/05.  
 PT New polynucleotides encoding human secreted proteins - derived from  
 PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,  
 PT ovary, pituitary, retina and colon cDNA libraries  
 PS Claim 1; Page 341; 64pp; English.  
 CC The present sequence represents an expressed sequence tag (EST), and is  
 CC a polynucleotide of the invention. The polynucleotides of the invention  
 CC are all secreted EST sequences isolated from a variety of human tissue  
 CC sources. The EST sequences and proteins encoded by them are predicted to  
 CC have useful biological activities which would make them suitable for  
 CC treating, preventing or ameliorating medical conditions in humans and  
 CC animals, although no supporting data is given. Suggested activities  
 CC include nutritional activity, immune stimulating or suppressing activity,  
 CC haematopoiesis regulating activity, tissue growth activity, haemostatic  
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory  
 CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition  
 CC activity. The EST sequences are also stated to be useful for gene  
 CC therapy.  
 SQ Sequence 266 BP; 50 A; 56 C; 114 G; 46 T;

OM of: US-09-135-238b-2\_COPY\_273\_390 to: N\_Geneseq\_36:\* out\_format : pfs

Date: Sep 12, 2000 6:56 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

#### Command line parameters:

```
-MODEL=frame+p2n.model -DEV=xlp
-O/cgnt2.1/USPTO.spool/US09135238/runat_24082000_091324_17918/app_query.fasta_1.1282
-Da-N_Geneseq_36 -OPMn=fastap -SUFFIX=p2n.rng -GAPOP=12,000
-DELOP=6,000 -MNMATCH=0,100 -LOAPCL=0,000 -LOPEXT=0,000
-DELOP=6,000 -MNMATCH=0,100 -LOAPCL=0,000 -LOPEXT=0,000
-DELOP=6,000 -MNMATCH=0,100 -LOAPCL=0,000 -LOPEXT=0,000
-DELOP=6,000 -MNMATCH=0,100 -LOAPCL=0,000 -LOPEXT=0,000
-DELOP=6,000 -MNMATCH=0,100 -LOAPCL=0,000 -LOPEXT=0,000
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM-ext -MINLEN=0 -MAXLEN=1000000
-USER=US09135238_cgnt1_1.123 -NCPV=6 -ICPV=3 -LONGLOS -NO_XLPXY
-WAIT -THREADS=1
```

#### Search information block:

```
Query: US-09-135-238b-2_COPY_273_390
Query length: 118
Database: N_Geneseq_36:*
Database sequences: 311585
Database length: 125096042
Search time (sec): 121.040000
```

#### score\_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
N_Geneseq_36:X28178	+	616.00	734.99	2040	Human PIGRL-1 coding sequence.
N_Geneseq_36:T93610	+	101.50	114.54	88.60	Mycobacterium tuberculosis gen
N_Geneseq_36:T93610	+	100.00	133.41	7.87	EST clone EAL10. New polynucle
N_Geneseq_36:T93610	+	100.00	118.33	54.45	Encodes human Retinoic Acid R
N_Geneseq_36:T93610	+	100.00	118.33	54.45	Human retinoic acid receptor g
N_Geneseq_36:T93610	+	96.50	115.09	82.56	Human dmpase gene 5' region.
N_Geneseq_36:T93610	+	96.00	117.41	61.31	31-O-desmethyl-FK506 O-methyl
N_Geneseq_36:T93610	+	96.00	113.27	104.27	Nucleotide sequence of human
N_Geneseq_36:T93610	+	95.50	113.62	99.62	hNET cDNA sequence from p1 cld
N_Geneseq_36:T93610	+	95.00	113.62	99.62	cDNA encoding human netrin. Hu
N_Geneseq_36:T93610	+	95.00	92.81	1.4e+03	Streptomyces roseofulvus fren
N_Geneseq_36:T93610	+	95.00	92.81	1.4e+03	Streptomyces roseofulvus fren
N_Geneseq_36:T93610	+	94.50	113.37	102.96	Human Sox1 encoding cDNA. New cl
N_Geneseq_36:T93610	+	94.00	119.54	46.63	hRAR gamma A clone cDNA. New cl
N_Geneseq_36:T93610	+	94.00	87.02	3.0e+03	Platenolide synthase gene clu
N_Geneseq_36:T93610	+	94.00	87.02	3.0e+03	Platenolide synthase gene clu
N_Geneseq_36:T93610	+	93.50	115.57	77.65	Mouse CAAAX processing enzyme R
N_Geneseq_36:T93610	+	93.50	114.51	88.89	Mouse CAAAX processing enzyme R
N_Geneseq_36:T93610	+	93.00	107.20	227.06	S. avermitilis bkd gene region
N_Geneseq_36:T93610	+	92.00	122.15	33.39	Human BPS2 repressor factor (EP
N_Geneseq_36:T93610	+	91.50	110.41	150.54	Human CAAAX processing enzyme R
N_Geneseq_36:T93610	+	91.50	109.74	163.86	Human CAAAX processing enzyme R
N_Geneseq_36:T93610	+	91.50	109.72	164.28	Human RCBI (hRCBI) polypeptide
N_Geneseq_36:T93610	+	91.50	89.88	2.1e+03	Polynucleotide sequence from
N_Geneseq_36:T93610	+	91.00	110.18	185.01	Streptomyces lividans XP55 gen
N_Geneseq_36:T93610	+	91.00	108.76	185.90	Alcaligenes bronchisepticus ar
N_Geneseq_36:T93610	+	91.00	104.30	329.59	Genomic DNA encoding human CNE
N_Geneseq_36:T93610	+	91.00	101.40	478.02	DNA of hspEco-1 encoding the
N_Geneseq_36:T93610	+	91.00	101.40	478.02	Human C-type natriuretic pepti
N_Geneseq_36:T93610	+	90.50	106.59	245.45	Sequence encoding high density
N_Geneseq_36:T93610	+	90.50	103.36	286.67	HSV-2 strain 885 Contig ID 11
N_Geneseq_36:T93610	+	90.50	104.31	329.11	HSV-2 strain 885 Contig ID 90
N_Geneseq_36:T93610	+	90.00	107.13	229.27	PRV 28k gene. Mutant pseudo-ra
N_Geneseq_36:T93610	+	90.00	86.83	3.1e+03	Streptomyces roseofulvus fren
N_Geneseq_36:T93610	+	90.00	86.83	3.1e+03	Streptomyces roseofulvus fren
N_Geneseq_36:T93610	+	89.50	102.09	437.09	Receptor type tyrosine kinase
N_Geneseq_36:T93610	+	89.50	100.63	527.17	Non-B, non-C, non-G hepatitis
N_Geneseq_36:T93610	+	89.50	99.95	575.16	Human adeno-associated virus
N_Geneseq_36:T93610	+	89.00	117.19	63.08	DNA fragment containing the am
N_Geneseq_36:T93610	+	89.00	114.79	85.81	DNA fragment of PAPO-G encodes

N\_Geneseq\_36:003224 + 89.00 111.55 129.91 828 ! DNA fragment of PAPO-SX enco  
N\_Geneseq\_36:003223 + 89.00 111.39 132.69 846 ! DNA fragment of PAPO-SX enco  
N\_Geneseq\_36:003220 + 89.00 110.95 140.39 896 ! DNA fragment of PAPO-GMCSF e  
N\_Geneseq\_36:003221 + 89.00 110.87 141.78 905 ! DNA fragment of PAPO-GMCSF e  
N\_Geneseq\_36:V11747 - 89.00 107.30 224.27 1443 ! Human MNTF-1-1443 DNA fragm

seq\_name: N\_Geneseq\_36:X28178

#### seq documentation block:

ID X28178 standard; cDNA; 2040 BP.  
AC X28178;  
AT 16-JUN-1999 (first entry)  
DE Human PIGRL-1 coding sequence.  
KW PIGRL-1; human; autoimmune disease; hyper-IgM Immunodeficiency; HIM;  
KW X-linked Severe Combined Immunodeficiency; XSCID; IgA deficiency;  
OS Homo sapiens.  
KW diagnosis; therapy; ss.  
PN EP-905238-A2.  
PD 31-MAR-1999.  
PF 14-AUG-1998; 306487.  
PR 30-OCT-1997; US-961564.  
PR 25-AUG-1997; US-056935.  
PA (SMIK ) SMITHKLINE BEECHAM CORP.  
PI Sweet RW, Truneh A, Wu S;  
DR WPI: 99-192666/17.  
DR P-PSDB: Y05001.  
PT New polypeptides encoding PIGRL-1 useful for treating diseases such  
as X-linked Severe Combined Immunodeficiency  
PS Claim 4; Page 18-19; 26pp; English  
CC This sequence encodes the human PIGRL-1 protein of the invention  
CC may include hyper-IgM immunodeficiency (HIM), X-linked Severe Combined  
CC immunodeficiency (XSCID) and IgA deficiency. These diseases can be  
CC diagnosed or susceptibility to them predicted by: (1) determining whether  
CC there is a mutation in the genomic copy of the gene encoding PIGRL-1; or  
CC (2) measuring the amount of PIGRL-1 in a sample derived from the patient.  
CC Patients deficient in PIGRL-1 can be treated by administering either the  
CC PIGRL-1 DNA or its complement or an agonist of PIGRL-1 to the patient.  
CC Patients with excessive expression or activity of PIGRL-1 can be treated  
CC by administering an antagonist of PIGRL-1, an antisense nucleic acid  
CC molecule which inhibits the expression of PIGRL-1 or administering  
CC sufficient PIGRL-1 to compete with the endogenous activity. PIGRL-1 can  
CC be used to identify its agonists by contacting a cell expressing PIGRL-1  
CC with a candidate compound in the presence of a signal system and noting  
CC the candidate as an agonist if a signal is produced. The same method can  
CC be used to identify antagonists of PIGRL-1 but the presence of an  
CC antagonist is indicated by a decrease in production of the signal.  
CC Antibodies against PIGRL-1 may be used to isolate or identify clones  
CC expressing PIGRL-1. Polynucleotides encoding PIGRL-1 may be used to  
CC identify chromosomal mutations in the gene encoding PIGRL-1 in patients.  
CC This information may then be correlated with the incidence of autoimmune  
CC disease in those patients to identify whether the mutation causes the  
CC disease.

Sequence 2040 BP; 478 A; 616 C; 494 G; 452 T;

#### alignment\_scores:

Quality: 616.00 Length: 118  
Ratio: 5.220 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

#### alignment\_block:

US-09-135-238b-2\_COPY\_273\_390 x X28178 ..  
Align seg 1/1 to: X28178 from: 1 to: 2040

1 LysArgAlaValcUArgArgLysAlaLeuSerArgArgAlaArgLe 17  
962 AAAAGGGCGCTTGAAGGAGGAAGAGCTCTCCAGCGCGCCGCCGAC 1011  
17 uAlaValArgMetArgAlaLeuGluSerGlnArgProArgGlySerp 34  
1012 GCCCGTGAGGATGCGCGCCCTGGAGAGCTCCAGAGCCCGCGCGTGC 1061



Db 423 GCTAAAGG--AATTACTTGATGATGATAAACAACTTATTANAGATATNATATATA 479

Search completed: September 12, 2000, 13:57:06  
Job time: 13667 sec

High quality sequence stop: 390.

## FEATURES

Location/Qualifiers  
 1. .404  
 /organism="Homo sapiens"  
 /db\_xref="GDB:5044084"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:813174"  
 /tissue\_type="Pooled human melanocyte, fetal heart, and pregnant uterus"  
 /lab\_host="DH10B"  
 /note="Organ: mixed (see below); Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (melanocyte 2N8HM, pregnant uterus N8HP0, and fetal heart N8H19W) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 260232-265223, 340488-345479, and 484488-489479."  
 71 a 153 c 114 g 66 t

BASE COUNT

ORIGIN

Query Match 19.9%; Score 380; DB 26; Length 404;  
 Best Local Similarity 99.5%; Pred. No. 1.2e-93;  
 Matches 402; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 912 aagccctctccaggcgcccgactgcccgtgaggtgagcgccctggagagctccc 971  
 Db 2 AAGCCCTCTCCAGCGCGCGCGGCGGCTGAGGATGGCGCCCTGGAGAGCTCCC 61

QY 972 agagcccgccggtccgcgacgcgcctcccaaaacatctacagcgctccgcgc 1031  
 Db 62 AGAGCGCCCGCGGCTGCCG-GACCGCCCTCCCAACACATCTACAGCGCTCCCGC 120

QY 1032 ggcgcgctc-tggagcgagcgtgcaggcacagggaggcccgctcccgcccgag 1090  
 Db 121 GGGCGCGCTGCTGGAGCGGACGCTGGAGCGACAGGGAGCGCCCGCTTCCGCGCGCGAG 180

QY 1091 cgcgctgcccccgcccgctgcaggtgctgaatctccctggctccatgcccatctc 1150  
 Db 181 CGCCGTTGCGCCCGCGCGCTGCAGGTGCTGAATCTCCCTGGCTCCATGCCCATCTC 240

QY 1151 tgaagaccagctgtgaatacgtgagcctctaccacagcgtgcgcgcatgatggagaca 1210  
 Db 241 TGAAGACCAAGCTGTGAATAGTGAGCGCTTACCACAGCGCTGCCGCCATGATGGAGACA 300

QY 1211 gtgattcagatgactacaatgttccctgctgcagacacaccccgagctatcccccaacc 1270  
 Db 301 GTGATTTCAGATGACTACATCAATGTCTCTGCTGACAACTCCCGCAGCTATCCCGCAACC 360

QY 1271 cagcgtcagactgtgtgccaagagctctcatctatctgctaat 1314  
 Db 361 CAGCCTCGACTGTGTGTCACAGGAGTCTCATCTCTCTGAT 404

RESULT 15

W01276

LOCUS

DEFINITION za04012.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:295055 5', mRNA sequence. 18-APR-1996

ACCESSION

W01276

VERSION

W01276.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 528)

AUTHORS

Hillier, L., Clark, N., Dubouque, T., Elliston, K., Hawkins, M.,

TITLE  
JOURNAL  
COMMENT

Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Maier, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P., and Willson, R.  
 The WashU-Merck EST Project  
 Unpublished (1995)  
 On Jan 7, 1998 this sequence version replaced gi:947771.  
 Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.edu  
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: E1primer  
 High quality sequence stop: 451.

FEATURES  
source

Location/Qualifiers  
 1. .528  
 /organism="Homo sapiens"  
 /db\_xref="GDB:123976"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:295055"  
 /clone\_lib="Soares fetal liver spleen 1NFLS"  
 /sex="male"  
 /dev\_stage="20 week-post conception fetus"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia) with a modified polylinker; Site 1: Pac I; Site 2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer 15', AAGTGAAGAATAAATAAGATCTTTTTTTTTTTTTTTT 3', double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."  
 129 a 125 c 119 g 149 t 6 others

BASE COUNT  
ORIGIN

Query Match 19.6%; Score 374; DB 90; Length 528;  
 Best Local Similarity 91.6%; Pred. No. 5.8e-92;  
 Matches 439; Conservative 0; Mismatches 33; Indels 7; Gaps 4;

QY 1396 gcctgagcatgctgtcccccagctgtcttcacacaccttgagccccctatgtg 1455  
 Db 6 GCCCTGAGCATGGCTGTGCCCGCCAGGCTGCTTGGACACACCTTGGCAGCCCCCTGTAGTTG 65

QY 1456 acaggttaagctgtagcctatagagcaattgtcccaatgcccactgtctcttcccaagc 1515  
 Db 66 ACAGGTAAGCTGTAGGCATGTAGAGCAATTGTCCCAATGCCACTGTCTCTTCCCAAGC 125

QY 1516 cgtcgacagactgtggagatttcagaggtttcttccatgtctttgaccacaggggtt 1575  
 Db 126 CGTCGACAGACTGTGGGATTTGCAGAGTGTTCTCCATGTCTTTGACCACAGGCT-TG 184

QY 1576 gttgtgcccaggtcttagatcatcagctcaggtcggggcagagcagctattgtct 1635  
 Db 185 TTGCTCCCGACGACTCTAGATCAGATGGCATCAGCTGGGGCAGAGGCATAGCTATTGTCT 244

QY 1636 cgggcatcttccccaggggtgggtcttaccacaaatagaaggctcttgcctgagttatgt 1695  
 Db 245 CGGGCATCTCTCCAGGGTGGGTCTTTACACAAATAGAGGCTTCTGTCTGTAGTTATGT 304

QY 1696 gaagtgcctcaagcccatgactaagcaggggtctgtat--aaacactcagaacgc 1753  
 Db 305 GACATGCTCAGCCCCCATGAGCTAGCAGGGGTCTGTTATATAAACAATCTCTGGAACGC 364

QY 1754 ctttgccctgatacccaattgttagcactgttagtgaacgtctacttctcaattctat 1813  
 Db 365 CTTTGGCCCTGATCCCAAAATGTTAGCACTTGTCTAGTAGTAACGCTCTACTTA--NTCAAGTCTAT 422

QY 1814 gctaaaggcaatttatctctgtgtgatgataaacaacattatttagcaagatagcata 1872

seq\_documentation\_block:  
AC  
AD  
AE  
AF  
AG  
AH  
AI  
AJ  
AK  
AL  
AM  
AN  
AO  
AP  
AQ  
AR  
AS  
AT  
AU  
AV  
AW  
AX  
AY  
AZ  
BA  
BB  
BC  
BD  
BE  
BF  
BG  
BH  
BI  
BJ  
BK  
BL  
BM  
BN  
BO  
BP  
BQ  
BR  
BS  
BT  
BU  
BV  
BW  
BX  
BY  
BZ  
CA  
CB  
CC  
CD  
CE  
CF  
CG  
CH  
CI  
CJ  
CK  
CL  
CM  
CN  
CO  
CP  
CQ  
CR  
CS  
CT  
CU  
CV  
CW  
CX  
CY  
CZ  
DA  
DB  
DC  
DD  
DE  
DF  
DG  
DH  
DI  
DJ  
DK  
DL  
DM  
DN  
DO  
DP  
DQ  
DR  
DS  
DT  
DU  
DV  
DW  
DX  
DY  
DZ  
EA  
EB  
EC  
ED  
EE  
EF  
EG  
EH  
EI  
EJ  
EK  
EL  
EM  
EN  
EO  
EP  
EQ  
ER  
ES  
ET  
EU  
EV  
EW  
EX  
EY  
EZ  
FA  
FB  
FC  
FD  
FE  
FF  
FG  
FH  
FI  
FJ  
FK  
FL  
FM  
FN  
FO  
FP  
FQ  
FR  
FS  
FT  
FU  
FV  
FW  
FX  
FY  
FZ  
GA  
GB  
GC  
GD  
GE  
GF  
GG  
GH  
GI  
GJ  
GK  
GL  
GM  
GN  
GO  
GP  
GQ  
GR  
GS  
GT  
GU  
GV  
GW  
GX  
GY  
GZ  
HA  
HB  
HC  
HD  
HE  
HF  
HG  
HH  
HI  
HJ  
HK  
HL  
HM  
HN  
HO  
HP  
HQ  
HR  
HS  
HT  
HU  
HV  
HW  
HX  
HY  
HZ  
IA  
IB  
IC  
ID  
IE  
IF  
IG  
IH  
II  
IJ  
IK  
IL  
IM  
IN  
IO  
IP  
IQ  
IR  
IS  
IT  
IU  
IV  
IW  
IX  
IY  
IZ  
JA  
JB  
JC  
JD  
JE  
JF  
JG  
JH  
JI  
JJ  
JK  
JL  
JM  
JN  
JO  
JP  
JQ  
JR  
JS  
JT  
JU  
JV  
JW  
JX  
JY  
JZ  
KA  
KB  
KC  
KD  
KE  
KF  
KG  
KH  
KI  
KJ  
KK  
KL  
KM  
KN  
KO  
KP  
KQ  
KR  
KS  
KT  
KU  
KV  
KW  
KX  
KY  
KZ  
LA  
LB  
LC  
LD  
LE  
LF  
LG  
LH  
LI  
LJ  
LK  
LL  
LM  
LN  
LO  
LP  
LQ  
LR  
LS  
LT  
LU  
LV  
LW  
LX  
LY  
LZ  
MA  
MB  
MC  
MD  
ME  
MF  
MG  
MH  
MI  
MJ  
MK  
ML  
MM  
MN  
MO  
MP  
MQ  
MR  
MS  
MT  
MU  
MV  
MW  
MX  
MY  
MZ  
NA  
NB  
NC  
ND  
NE  
NF  
NG  
NH  
NI  
NJ  
NK  
NL  
NM  
NN  
NO  
NP  
NQ  
NR  
NS  
NT  
NU  
NV  
NW  
NX  
NY  
NZ  
OA  
OB  
OC  
OD  
OE  
OF  
OG  
OH  
OI  
OJ  
OK  
OL  
OM  
ON  
OO  
OP  
OQ  
OR  
OS  
OT  
OU  
OV  
OW  
OX  
OY  
OZ  
PA  
PB  
PC  
PD  
PE  
PF  
PG  
PH  
PI  
PJ  
PK  
PL  
PM  
PN  
PO  
PP  
PQ  
PR  
PS  
PT  
PU  
PV  
PW  
PX  
PY  
PZ  
QA  
QB  
QC  
QD  
QE  
QF  
QG  
QH  
QI  
QJ  
QK  
QL  
QM  
QN  
QO  
QP  
QQ  
QR  
QS  
QT  
QU  
QV  
QW  
QX  
QY  
QZ  
RA  
RB  
RC  
RD  
RE  
RF  
RG  
RH  
RI  
RJ  
RK  
RL  
RM  
RN  
RO  
RP  
RQ  
RR  
RS  
RT  
RU  
RV  
RW  
RX  
RY  
RZ  
SA  
SB  
SC  
SD  
SE  
SF  
SG  
SH  
SI  
SJ  
SK  
SL  
SM  
SN  
SO  
SP  
SQ  
SR  
SS  
ST  
SU  
SV  
SW  
SX  
SY  
SZ  
TA  
TB  
TC  
TD  
TE  
TF  
TG  
TH  
TI  
TJ  
TK  
TL  
TM  
TN  
TO  
TP  
TQ  
TR  
TS  
TT  
TU  
TV  
TW  
TX  
TY  
TZ  
UA  
UB  
UC  
UD  
UE  
UF  
UG  
UH  
UI  
UJ  
UK  
UL  
UM  
UN  
UO  
UP  
UQ  
UR  
US  
UT  
UU  
UV  
UW  
UX  
UY  
UZ  
VA  
VB  
VC  
VD  
VE  
VF  
VG  
VH  
VI  
VJ  
VK  
VL  
VM  
VN  
VO  
VP  
VQ  
VR  
VS  
VT  
VU  
VV  
VW  
VX  
VY  
VZ  
WA  
WB  
WC  
WD  
WE  
WF  
WG  
WH  
WI  
WJ  
WK  
WL  
WM  
WN  
WO  
WP  
WQ  
WR  
WS  
WT  
WU  
WV  
WW  
WX  
WY  
WZ  
XA  
XB  
XC  
XD  
XE  
XF  
XG  
XH  
XI  
XJ  
XK  
XL  
XM  
XN  
XO  
XP  
XQ  
XR  
XS  
XT  
XU  
XV  
XW  
XX  
XY  
XZ  
YA  
YB  
YC  
YD  
YE  
YF  
YG  
YH  
YI  
YJ  
YK  
YL  
YM  
YN  
YO  
YP  
YQ  
YR  
YS  
YT  
YU  
YV  
YW  
YX  
YY  
YZ  
ZA  
ZB  
ZC  
ZD  
ZE  
ZF  
ZG  
ZH  
ZI  
ZJ  
ZK  
ZL  
ZM  
ZN  
ZO  
ZP  
ZQ  
ZR  
ZS  
ZT  
ZU  
ZV  
ZW  
ZX  
ZY  
ZZ



```

22 ArgAlaLeuGluSerSerGlnArgProArglySerProArg..... 35
   ::::| | | | | | | | | | | | | | | | | | | | | |
1020 AAGCGCGCCCTCGCCAGGGGCCCCCTTAATGACACCCACCCCGCCG 971
   36 . ProArgSerGlnAsnAsn..... IleTyrSerAlaCysProA 48
   | | | | | | | | | | | | | | | | | | | | | |
970 CCCCCCGCGGAGAACACACCGCTCGCACCTACTCTCCACAGCTCA 921
   48 rgrArgAlaArgGlyAlaAspAlaAlaGlyThrGlyAlaPro... Val 63
   | | | | | | | | | | | | | | | | | | | | | |
920 GCCGCTGATTCCTTCCAGAGCGCGGAGCCCTCCCGCCCTTCTCTA 871
   64 ProGlyProGlyAlaProLeuProProAla..... ProLeuGly 76
   | | | | | | | | | | | | | | | | | | | | | |
870 CCGCGCGCGCGCGCGGCTTCCCGCCCTTGAAGACCTCGCCCTTCA 821
   76 n..... ValSerGluSerProTyrLeuHisAlaP 86
   | | | | | | | | | | | | | | | | | | | | | |
820 GCCAGCGCGCGCGACTGACTGCTCCGCGCAGCTTGCTCAGCGCGC 771
   86 roserleu 88
   | | | | | | | | | | | | | | | | | | | | | |
770 CACCGCTG 763
seq_name: N_Geneseq_36:T03650

```

```

seq_documentation_block:
ID T03650 standard; DNA: 1150 BP.
AC T03650:
DE 16-AUG-1996 (first entry)
KW 31-O-desmethyl-FK506 O:methyltransferase gene FKMT2.
KW Streptomycetes; 31-O-desmethyl-FK506 O:methyltransferase; terminator;
KW expression cassette; strong promoter; PermE*; ribosome binding site;
KW translation start site; transcription terminator; selection marker;
KW hygromycin integration element; multicircle; chromosome; E. coli;
KW replicon; ColEI; overexpression; immunosuppressant drug; ss.
OS Streptomycetes sp. strain MA6548.
FH Key Location/Qualifiers
   cds
   FT 62..844
   FT /tag= a
   FT /product= 31-O-desmethyl-FK506 O:methyltransferase
   FT 842..893
   FT repeat_unit
   FT /tag= b
   FT /note= "inverted repeat unit"
   FT 845..877
   FT repeat_unit
   FT /tag= c
   FT /note= "inverted repeat unit"
   FT
   PN W09600282-A1.
   PD 04-JAN-1996.
   PF 19-JUN-1995; U07784.
   PR 23-JUN-1994; US-264861.
   PA (MERI ) MERCK & CO INC.
   PI Motamedi H, Shafiee A;
   DR WPI: 96-068866/07
   PT Integrative and replicative expression vectors for Streptomycetes
   PT esp. for expressing 31-O-desmethyl-FK506 O:methyltransferase for
   PT prodn of the immunosuppressive drug FK506
   PS Claim 23; Page 36; 48pp; English.
   CC This is the nucleotide sequence of the Streptomycetes gene FKMT2 encoding
   CC 31-O-desmethyl-FK506 O:methyltransferase. The sequence contains the
   CC coding region for the protein and includes the termination sequence.
   CC The gene was used in the construction of a Streptomycetes expression
   CC cassette comprising: (a) a strong promoter e.g. the PermE* constitutive
   CC promoter, (b) a synthetic ribosome binding site, (c) a translation start
   CC site (contained within the restriction enzyme site for NdeI), (d) a
   CC multiple cloning site, (e) a transcription terminator downstream from the
   CC MCS, esp. the terminator region of the FKMT2 gene. The vector also
   CC contains a selection marker, pref. the hygromycin resistance gene, an
   CC integration element, pref. from a multicircle, for integration into the
   CC Streptomycetes chromosome and an E.coli replicon, pref. the ColEI replicon.
   CC The vectors can be used for the integration of single copies of desired
   CC genes into the Streptomycetes genome or for the transfer of multiple copies

```

CC of genes of interest for overexpression purposes, e.g. for the conversion  
 CC of 31-desmethyl-FK506 to the immunosuppressant drug FK506 by integration  
 CC of the FKMT2 gene.  
 SO Sequence 1150 BP; 164 A; 418 C; 396 G; 172 T;

alignment\_scores:  
 Quality: 96.00 Length: 137  
 Ratio: 1.333 Gaps: 6  
 Percent Similarity: 52.555 Percent Identity: 29.927

alignment\_block:  
 US-09-135-238b-2\_copy\_273\_390 x T03650 ..

Align seg 1/1 to: T03650 from: 1 to: 1150

```

1 LysArgAlaValAlaGluArgGlyAlaLeuSerArgArgAlaArgGly 17
   ::::| | | | | | | | | | | | | | | | | | | | | |
697 CGTCCGCGAGGTCACAGACGTGAGGGCGCGCTCGACAGAGTCTGCGC 746
   17 uAlaVal..... ArgMetArgAlaLeuGluSerSerGln 29
   | | | | | | | | | | | | | | | | | | | | | |
747 TGCTGCGCGCGCCATGCTTCACCGCTGCTGCCGAGCAGATCCGCTTTC 796
   29 rgrProArglySerProArgProArgSerGlnAsnAsnIleTyrSerAla 45
   | | | | | | | | | | | | | | | | | | | | | |
797 GCGGCGACGACATCCACAGGTCCGCC..... ACGGCG 831
   46 CysProArgArgAlaArgGlyAlaAspAlaAlaGlyThrGly..... 59
   | | | | | | | | | | | | | | | | | | | | | |
832 TCGGACCCACTGAGCGCGCGCGCGGCTACCCGACCGCGCGCGCG 881
   60 ..... GlnAlaProValPro..... GlyProGlyAlaP 69
   | | | | | | | | | | | | | | | | | | | | | |
882 TCGCGCGGCTCGCCGACGTCGCGCGCGCTCTTCGGAACCTGCTGGCG 931
   69 roLeuProProAlaProLeuGlnValSerGluSerProTyrLeuHisAla 85
   | | | | | | | | | | | | | | | | | | | | | |
932 CTCACCCGCCAGCTTTCGGGAACAGTTCGTAGGTGCTGCCACCGTG 981
   86 ProSerLeuLysThrSerCysGluTyrValSerLeuTyrHisGlnPro.. 101
   | | | | | | | | | | | | | | | | | | | | | |
982 CTGCGCGCTGACGACAGACTGCTCGGCGATCTCCGTGTGACGCCGAC 1031
   102 ..... AlaAlaMetMetGluAspSerAspSer... AspAspTyrIleA 115
   | | | | | | | | | | | | | | | | | | | | | |
1032 CGGCGCCAGCGCGGCCACCGCGCGCTCCAGCTCCGTCAGTGCAGCGATCC 1081
   115 snValProAla 118
   | | | | | | | | | | | | | | | | | | | | | |
1082 GCTGCCCGCAGC 1092
seq_name: N_Geneseq_36:V16346
seq_documentation_block:
ID V16346 standard; CDNA: 1974 BP.
AC V16346:
DE 03-JUN-1998 (first entry)
KW Nucleotide sequence of human netrin cDNA including 5' and 3' UTRs.
KW Human; netrin; hnMT; ATP binding cassette transporter; hABC3;
KW ribosomal L3; RPL3L; augmeneter of liver regeneration; hALR; treatment;
KW trapping; modulation; expression; antibody; identification; binding;
KW chemottractant; spinal commissural axon; neural regeneration;
KW axon growth; orientation; substrate specificity; ligand; exon trap; ss.
OS Homo sapiens.
PN W09748797-A1.
PD 24-DEC-1997.
PF 16-JAN-1997; U00785.
PR 09-DEC-1996; US-762500.
PR 17-JUN-1996; US-665259.
PR 01-OCT-1996; US-720614.
PA (GENZ ) GENZYME CORP.
PI Burn TC, Connors TD, Dackowski WR, Klingner KW, Landes GW.

```



```

      : ::::::::::::::::::::::::::::::
1521 CACCACAGCGCTACTCCGCGCGCGCTCTCTCCGCTCCGGA 1472
      : ::::::::::::::::::::::::::::::
57 1yThrgly.....GluAlaProValProgly 65
      : ::::::::::::::::::::::::::::::
1471 ACACGGCGAGACCGCCACCGGAGAGGTGTCACGCGCGCGCGCTCG 1422
      : ::::::::::::::::::::::::::::::
66 ProGlyAlaProLeu.ProProAlaProLeuGlnValSerGluSerProt 82
      : ::::::::::::::::::::::::::::::
1421 CCGCGCGACCCACCGCCACCTGCACCG..... 1394
      : ::::::::::::::::::::::::::::::
82 rpleuHisAlaProSerLeuLysThr.....SerCys 92
      : ::::::::::::::::::::::::::::::
1393 .....CATAGTCTTCTTGCAAGAACTTTAGCGCTATGCGGTAAGTGC 1349
      : ::::::::::::::::::::::::::::::
93 .....GluTyValSerLeuTyHisGlnProAlaAla 103
      : ::::::::::::::::::::::::::::::
1348 CACGGGCAAGTTTGACAGTCGACAGTCAGTCTGGGCT 1310
      : ::::::::::::::::::::::::::::::

seq_name: N_Geneseq_36:V16306
seq_documentation_block:
ID V16306 standard; cDNA; 1743 BP.
AC V16306;
DT 03-JUN-1998 (first entry)
DE cDNA encoding human netrin.
KW Human; netrin; hNET; ATP binding cassette transporter; hABC3;
KW ribosomal L3; RPL3L; augment of liver regeneration; hALR; treatment;
KW trapping; modulation; expression; antibody; identification; binding;
KW chemottractant; spinal commissural axon; neural regeneration;
KW axon growth; orientation; substrate specificity; ligand; exon trap; ss.
OS Homo sapiens.
FH Key 1.1743
FT CDS /product= netrin
FT
FT
FN W09748797-A1.
PD 24-DEC-1997.
PR 16-JAN-1997; U00785.
PR 09-DEC-1996; US-762500.
PR 17-JUN-1996; US-665259.
PR 01-OCT-1996; US-720614.
PA (GENZ ) GENZYME CORP.
PI Burn TC, Connors TD, Dackowski WR, Klinger KW, Landes GM.
PI Van Raay TJ;
PI WPI: 98-063138/06.
DR P-PSDB: W46758.
DE Human chromosome 16 genes encoding netrin, ATP binding cassette
PT transporter, ribosomal L3 and augment of liver regeneration
PT proteins - useful for, e.g. treatment of liver disease and cystic
PT fibrosis
PS Claim 4; Fig 4A-B; 220bp; English.
CC The present sequence encodes human netrin (hNET). Netrins define a family
CC of chemotropic factors which have been shown to play a central role in
CC axon guidance. GRAL12 analysis predicts 6 exons within the genomic DNA
CC sequence, with 5 exons encoding sequences with homology to chicken
CC netrins. Chicken netrins have been shown to function as chemottractants
CC for developing spinal commissural axons. Human netrins may therefore have
CC a significant role in neural regeneration. Though netrins do not by
CC themselves promote axon growth, they do play a role in the orientation of
CC axon growth. The sequence was isolated using an exon trap. Sequences
CC encoding human ATP binding cassette transporter (hABC3), human ribosomal
CC L3 (RPL3L), and human augment of liver regeneration (hALR) were also
CC isolated. The antisense oligonucleotides of the present sequence are used
CC to modulate expression of hNET prevent its translation. Antibodies
CC against hNET can be used to block binding of its naturally occurring
CC ligands. Host cells containing vectors with DNA inserts encoding the
CC protein can be used in a method for identifying compounds which bind
CC to hNET.
SQ Sequence 1743 BP; 231 A; 652 C; 589 G; 271 T;

```

```

      Quality: 95.50      Length: 147
      Ratio: 1.540      Gaps: 8
Percent Similarity: 42.177      Percent Identity: 27.211

alignment_block:
us-09-135-238b-2_copy_273_390 x V16306/rev ..
Align seg 1/1 to reverse of: V16306 from: 1 to: 1743

12 ArgArgAlaArgArgLeuAlaValArgMetArgAlaLeuGluSerSer.. 27
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1721 CGCGTTCGCGTCTGTCAGAGCTCCGACAGCGCCCGCGTCCAGCGTCCCT 1672
27 .....
27 .....
1671 CCAGGTCAGCAGAGGCTTCGCGGCGCGCGATGAGCCGGGCCCCGGC 1622
27 .....
1621 CCCCCGGCCCCCAGCGCGGCTCCAGGCCGCCGCCAGAGAGTAG 1572
28 GlnArgPro.....ArgGlySerProArgProArgSerGlnAsnAs 41
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1571 CGCGCGCGCGGAGACAGCGCGCGGACGCGAGCGCGCATCCCGCGGG 1522
41 nileTySerAla...CysProArgArgAlaArgGlyAlaAspAlaIag 57
: ::::::::::::::::::::::::::::::
1521 CACCACAGCGCGCTACTCCGCGCGCGCTCTCTCCGCTCCGGA 1472
      : ::::::::::::::::::::::::::::::
57 1yThrgly.....GluAlaProValProgly 65
      : ::::::::::::::::::::::::::::::
1471 ACACGGGACACCGCCACCGGAGAGGTGTCACGCGCGCGCGCTCG 1422
      : ::::::::::::::::::::::::::::::
66 ProGlyAlaProLeu.ProProAlaProLeuGlnValSerGluSerProt 82
      : ::::::::::::::::::::::::::::::
1421 CCGCGCGCACCCACCGCCACCTGCACCG..... 1394
      : ::::::::::::::::::::::::::::::
82 rpleuHisAlaProSerLeuLysThr.....SerCys 92
      : ::::::::::::::::::::::::::::::
1393 .....CATAGTCTTCTTGCAAGAACTTTAGCGCTATGCGGTAAGTGC 1349
      : ::::::::::::::::::::::::::::::
93 .....GluTyValSerLeuTyHisGlnProAlaAla 103
      : ::::::::::::::::::::::::::::::
1348 CACGGGCAAGTTTGACAGTCGACAGTCAGTCTGGGCT 1310
      : ::::::::::::::::::::::::::::::

seq_name: N_Geneseq_36:T93095
seq_documentation_block:
ID T93095 standard; cDNA; 24379 BP.
AC T93095;
DT 11-MAY-1998 (first entry)
DE Streptomyces frenolicin gene cluster.
KW frenolicin; antibiotic; feed additive; anticoccidial;
KW coccidiostatic; efflux pump; butyrate starter synthase;
KW polyketide synthase; PKS; hemiketalase; ketoreductase; cyclase;
KW dehydrase; ketoreductase; hydroxylase; Streptomyces roseofulvus;
KW ds.
OS Streptomyces sp.
FH Key 1.2948
FT CDS /product= 80 kDa non-membrane protein
FT /note= "gene A (specifically claimed)"
FT 2945..3916
FT /tag= b
FT /product= membrane protein
FT /note= "gene B (specifically claimed)"
FT 4020..4844
FT /tag= c
FT /product= protein with 6 membrane-spanning domains
FT /note= "gene C (specifically claimed)"
FT 4841..6415
FT CDS /tag= d

```

alignment\_scores:

FT /product= ATP-binding component of ABC transporter  
FT /note= "gene D (specifically claimed)"  
FT CDS  
FT /tag= e  
FT /product= unknown non-membrane protein  
FT /note= "gene E (specifically claimed)"  
FT CDS  
FT /tag= f  
FT /product= putative efflux pump  
FT /note= "gene F (specifically claimed)"  
FT CDS  
FT /tag= g  
FT /product= transcription activator  
FT /note= "gene G (specifically claimed)"  
FT CDS  
FT /tag= h  
FT /product= translationally coupled to gene I  
FT /note= "gene H (specifically claimed)"  
FT CDS  
FT /tag= i  
FT /product= homologue of fabH  
FT /note= "gene I (specifically claimed)"  
FT CDS  
FT /tag= j  
FT /product= acyl carrier protein  
FT /note= "gene J (specifically claimed)"  
FT CDS  
FT /tag= k  
FT /product= acyltransferase  
FT /note= "gene K (specifically claimed)"  
FT CDS  
FT /tag= l  
FT /product= PKS ketosyl synthase subunit  
FT /note= "gene L (specifically claimed)"  
FT CDS  
FT /tag= m  
FT /product= PKS chain length factor  
FT /note= "gene M (specifically claimed)"  
FT CDS  
FT /tag= n  
FT /product= acyl carrier protein used by the PKS  
FT /note= "gene N (specifically claimed)"  
FT CDS  
FT /tag= o  
FT /product= putative hemiketal dehydrase  
FT /note= "gene O (specifically claimed)"  
FT CDS  
FT /tag= p  
FT /product= ketoreductase related to actIII  
FT /note= "gene P (specifically claimed)"  
FT CDS  
FT /tag= q  
FT /product= cyclase/dehydrase related to act VII  
FT /note= "gene Q (specifically claimed)"  
FT CDS  
FT /tag= r  
FT /product= cyclase/dehydrase related to act V  
FT /note= "gene R (specifically claimed)"  
FT CDS  
FT /tag= s  
FT /product= oxidoreductase  
FT /note= "gene S (specifically claimed)"  
FT CDS  
FT /tag= t  
FT /product= quinone-forming hydroxylase  
FT /note= "gene T (specifically claimed)"  
FT CDS  
FT /tag= u  
FT /product= not specified  
FT /note= "gene U"

EP-806480-A2.  
12-NOV-1997.  
PD 02-MAY-1997; 107329.

PR 07-MAY-1996; US-016753.  
PA (HOFF ) HOPFMAN LA ROCHE & CO AG F.  
PI Reeves CD, Soliday CL;  
DR WPI: 97-538619/50.  
DR P-PDB: W34199-W34219.  
PT Streptomycetes frenolicin gene cluster - useful for producing  
PT recombinant frenolicin antibiotics  
PS Claim 1; Page 40-60; 66pp; English.  
CC This DNA sequence comprises the Streptomycetes frenolicin gene  
CC cluster containing specifically claimed coding sequences (genes  
CC A-U) that respectively encode 21 proteins (see W34199-219) involved  
CC in frenolicin synthesis. The genes can be divided into 5  
CC subclusters: (1) genes A, B, C, D and E encode an efflux pump; (2)  
CC genes H, I, J and K encode butyrate starter synthases; (3) genes  
CC L, M and N encode polyketide synthases (PKS); (4) genes O, P, Q and  
CC R encode a hemiketalase, a ketoreductase and cyclases/dehydrases;  
CC and (5) genes S and T encode a keto/enoyl reductase and a  
CC hydrolase. Also claimed are vectors, host cells (especially a  
CC Streptomycetes sp., particularly streptomycetes roseofulvus), and the  
CC encoded proteins. Cells transformed using the above sequence can  
CC be cultured to produce frenolicins or frenolicin precursors. The  
CC precursors can be converted to frenolicins by chemical or other  
CC methods. The frenolicins can be oxidised to frenolicin B, an  
CC antibiotic used as an anticoccidial agent. The frenolicins can be  
CC used as animal feed additives.  
SQ Sequence 24379 BP; 3077 A; 9792 C; 8499 G; 3011 T;

alignment\_scores:  
Quality: 95.00 Length: 135  
Ratio: 1.439 Gaps: 9  
Percent Similarity: 48.889 Percent Identity: 32.593

## alignment block:

US-09-135-238B-2\_COPY\_273\_390 x T93095 ..

Align seg 1/1 to: T93095 from: 1 to: 24379

2 ATGAAValAlGluArgArgLysAlaLeuSerArgArgAlaArgLysLeu 18  
||||| ||||| |||  
21737 CGGACGAGTACGCGCGGTACCGCGCGGTACGCGAGAGAGCGCGCC 21786  
18 aValArgMetArgAlaLeuGluSer.....SerLArgProArgLys 33  
..... |||||  
21787 CACCGGCGCGAGGTGCAAGCCGTCGAGCGCAGAGAGCGCGTCGCGCAG 21836  
33 eProArg.....ProArgSerGlnAsn 41  
||||| |||||  
21837 TTCCCGAAGCGCGCTTCGCGCGCGCGCGCGCGCGCGCGCAAGTGC 21886  
42 IleArgSer.....AlaGysProArgArgAla..... 50  
||| |||||  
21887 GGGGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGCC 21936  
51 .....ArgGlyAlaAspAlaAlaGlyThrGlyGlu 61  
||||| |||||  
21937 CCGCTCTCTCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 21986  
61 IapProVal.....ProGlyProGlyAlaProLeuPro... 71  
||||| |||||  
21987 CCGCGCGCGGTTCGCGAGCGGTCCGAGG...GGCGTCCGCGCGCGTGC 22033  
72 ProAlaProLeuGlnValSerGluSerPro.....TripleuHi 84  
||||| |||||  
22034 CCGCGCGCGCGGTCCGCGCGGTGCGCGCGCGCGCGCGAGGCGTGCAT 22083  
84 sAlaProSerLeuArgThrSerGlyLysValSerLeuTyrHisGlnP 101  
||||| |||||  
22084 TGCCTCGTCA.....TCGCTAGCTACCTCCAGCGCAACATCTCC 22124  
101 roAla 102  
|||||  
22125 CGGCG 22129



```
seq_name: N_Geneseq_36.V25925
seq_documentation_block:
ID      V25925, standard; cDNA; 24379 BP.
AC      15-JUL-1998 (first entry)
DT      Streptomyces roseofulvus frenolicin gene cluster.
DE      Streptomyces roseofulvus; frenolicin gene cluster; frenolicin B.
KW      Streptomyces roseofulvus; ss.
OS      Streptomyces roseofulvus.
FH      location/Qualifiers
FH      Key
FH      CDS
FT      /tag= a
FT      /transl_except= (pos:636..638,aa:Met)
FT      /note= "encodes protein given in W55800"
FT      2945..3916
FT      CDS
FT      /tag= b
FT      /transl_except= (pos:2945..2947,aa:Met)
FT      /note= "encodes protein given in W55801"
FT      4020..4844
FT      CDS
FT      /tag= c
FT      /transl_except= (pos:4020..4022,aa:Met)
FT      /note= "encodes protein given in W55802"
FT      4841..6415
FT      CDS
FT      /tag= d
FT      /note= "encodes protein given in W55803"
FT      6533..7183
FT      CDS
FT      /tag= e
FT      /transl_except= (pos:6533..6535,aa:Met)
FT      /note= "encodes protein given in W55804"
FT      7344..8897
FT      CDS
FT      /tag= f
FT      /note= "encodes protein given in W55805"
FT      9164..10012
FT      CDS
FT      /tag= g
FT      /transl_except= (pos:9164..9166,aa:Met)
FT      /note= "encodes protein given in W55806"
FT      complement (10105..10621)
FT      CDS
FT      /tag= h
FT      /transl_except= (pos:10621..10619,aa:Met)
FT      /note= "encodes protein given in W55807"
FT      complement (10618..11628)
FT      CDS
FT      /tag= i
FT      /transl_except= (pos:11628..11626,aa:Met)
FT      /note= "encodes protein given in W55808"
FT      11809..12066
FT      CDS
FT      /tag= j
FT      /note= "encodes protein given in W55809"
FT      complement (12154..13209)
FT      CDS
FT      /tag= k
FT      /transl_except= (pos:13209..13207,aa:Met)
FT      /note= "encodes protein given in W55810"
FT      13409..14686
FT      CDS
FT      /tag= l
FT      /transl_except= (pos:13409..13411,aa:Met)
FT      /note= "encodes protein given in W55811"
FT      14767..16047
FT      CDS
FT      /tag= m
FT      /note= "encodes protein given in W55812"
FT      16120..16371
FT      CDS
FT      /tag= n
FT      /note= "encodes protein given in W55813"
FT      complement (16453..16935)
FT      CDS
FT      /tag= o
FT      /transl_except= (pos:16935..16933,aa:Met)
FT      /note= "encodes protein given in W55814"
FT      17088..18903
FT      CDS
FT      /tag= p
FT      /note= "encodes protein given in W55815"
FT      17993..18898
FT      CDS
FT      /tag= q
FT      /note= "encodes protein given in W55816"
FT
```

FT	CDS	18895..19839
FT	/*tag_x	/note="encodes protein given in W55817"
FT	complement (19907..20907)	
FT	/*tag_s	
FT	/transl_except- (pos:20907..20905,aa:Met)	
FT	/note="encodes protein given in W55818"	
FT	complement (20904..22094)	
FT	/*tag_t	
FT	/transl_except- (pos:22040..22038,aa:Arg)	
FT	/transl_except- (pos:22037..22035,aa:Arg)	
FT	/transl_except- (pos:22034..22032,aa:Thr)	
FT	/transl_except- (pos:22031..22029,aa:Ala)	
FT	/note="encodes protein given in W55819"	

PR Streptomycin gene cluster - useful as an antibiotic  
 PS Claim 2: Page 36-44; 50pp: Japanese.  
 CC The present sequence represents a frenolicin gene cluster from  
 CC *Streptomyces roseofulvus*. The present invention describes: (1) a  
 CC vector containing the frenolicin gene cluster DNA sequence operably  
 CC connected to an expression control sequence; (2) a host cell  
 CC transformed by the above vector; (3) a protein coded by the above DNA  
 CC sequence; (4) a method for the preparation of frenolicin or a  
 CC biosynthetic intermediate for it in which the above cell is cultured  
 CC and frenolicin or its biosynthetic intermediate is isolated from the  
 CC culture or the cell; (5) a method for the preparation of frenolicin B  
 CC by oxidizing frenolicin, and (6) a method for the preparation of a  
 CC feed composition by mixing frenolicin with other components.  
 CC Frenolicin B is useful as an antibiotic.  
 SQ Sequence 24319 BP; 3077 A; 9792 C; 8499 G; 3011 T;

```
alignment_scores:
  Quality: 95.00
  Ratio: 1.439
  Gaps: 9
Percent Similarity: 48.889
Percent Identity: 32.593
```

```
alignment_block:
US-09-135-238B-2_COPY_273_390 x V25925 ..
```

Align seg 1/1 to: V25925 from: 1 to: 24379

2 ArgAlaValGluArgArgLysAlaLeuSerArgArgAlaArgArgLeuAl 18

21737 CGGGCAGTACGGGCTAACGCCCGCGGTAGCGGAGAGCGAGCGGCC 21786

```
18 avaIargmetargalaleuciuser.....SergInargproargglys 33
```

21787 CACGCGCCGAGGTGCAGCCCGTCCCCGAGGCCGACGAGCGCCCTGGCGAG 21836

```

33 erProArg.....ProArgserGlnAsnAsn 41
      |||||      |||||:.....:

```

21837 TTCCCGAGCCGCTTCCGCGCGCGTGGGCGCGGGGCACGAGTGGC 21886

```

42 lleyrser.....alacysproatgargala..... 30
    ||          |||||
    ::

```

[illegible]

```

21 .....| 21
22 .....| 22
23 .....| 23
24 .....| 24
25 .....| 25
26 .....| 26
27 .....| 27
28 .....| 28
29 .....| 29
30 .....| 30
31 .....| 31
32 .....| 32
33 .....| 33
34 .....| 34
35 .....| 35
36 .....| 36
37 .....| 37
38 .....| 38
39 .....| 39
40 .....| 40
41 .....| 41
42 .....| 42
43 .....| 43
44 .....| 44
45 .....| 45
46 .....| 46
47 .....| 47
48 .....| 48
49 .....| 49
50 .....| 50
51 .....| 51
52 .....| 52
53 .....| 53
54 .....| 54
55 .....| 55
56 .....| 56
57 .....| 57
58 .....| 58
59 .....| 59
60 .....| 60
61 .....| 61
62 .....| 62
63 .....| 63
64 .....| 64
65 .....| 65
66 .....| 66
67 .....| 67
68 .....| 68
69 .....| 69
70 .....| 70
71 .....| 71
72 .....| 72
73 .....| 73
74 .....| 74
75 .....| 75
76 .....| 76
77 .....| 77
78 .....| 78
79 .....| 79
80 .....| 80
81 .....| 81
82 .....| 82
83 .....| 83
84 .....| 84
85 .....| 85
86 .....| 86
87 .....| 87
88 .....| 88
89 .....| 89
90 .....| 90
91 .....| 91
92 .....| 92
93 .....| 93
94 .....| 94
95 .....| 95
96 .....| 96
97 .....| 97
98 .....| 98
99 .....| 99
100 .....| 100
101 .....| 101
102 .....| 102
103 .....| 103
104 .....| 104
105 .....| 105
106 .....| 106
107 .....| 107
108 .....| 108
109 .....| 109
110 .....| 110
111 .....| 111
112 .....| 112
113 .....| 113
114 .....| 114
115 .....| 115
116 .....| 116
117 .....| 117
118 .....| 118
119 .....| 119
120 .....| 120
121 .....| 121
122 .....| 122
123 .....| 123
124 .....| 124
125 .....| 125
126 .....| 126
127 .....| 127
128 .....| 128
129 .....| 129
130 .....| 130
131 .....| 131
132 .....| 132
133 .....| 133
134 .....| 134
135 .....| 135
136 .....| 136
137 .....| 137
138 .....| 138
139 .....| 139
140 .....| 140
141 .....| 141
142 .....| 142
143 .....| 143
144 .....| 144
145 .....| 145
146 .....| 146
147 .....| 147
148 .....| 148
149 .....| 149
150 .....| 150
151 .....| 151
152 .....| 152
153 .....| 153
154 .....| 154
155 .....| 155
156 .....| 156
157 .....| 157
158 .....| 158
159 .....| 159
160 .....| 160
161 .....| 161
162 .....| 162
163 .....| 163
164 .....| 164
165 .....| 165
166 .....| 166
167 .....| 167
168 .....| 168
169 .....| 169
170 .....| 170
171 .....| 171
172 .....| 172
173 .....| 173
174 .....| 174
175 .....| 175
176 .....| 176
177 .....| 177
178 .....| 178
179 .....| 179
180 .....| 180
181 .....| 181
182 .....| 182
183 .....| 183
184 .....| 184
185 .....| 185
186 .....| 186
187 .....| 187
188 .....| 188
189 .....| 189
190 .....| 190
191 .....| 191
192 .....| 192
193 .....| 193
194 .....| 194
195 .....| 195
196 .....| 196
197 .....| 197
198 .....| 198
199 .....| 199
200 .....| 200
201 .....| 201
202 .....| 202
203 .....| 203
204 .....| 204
205 .....| 205
206 .....| 206
207 .....| 207
208 .....| 208
209 .....| 209
210 .....| 210
211 .....| 211
212 .....| 212
213 .....| 213
214 .....| 214
215 .....| 215
216 .....| 216
217 .....| 217
218 .....| 218
219 .....| 219
220 .....| 220
221 .....| 221
222 .....| 222
223 .....| 223
224 .....| 224
225 .....| 225
226 .....| 226
227 .....| 227
228 .....| 228
229 .....| 229
230 .....| 230
231 .....| 231
232 .....| 232
233 .....| 233
234 .....| 234
235 .....| 235
236 .....| 236
237 .....| 237
238 .....| 238
239 .....| 239
240 .....| 240
241 .....| 241
242 .....| 242
243 .....| 243
244 .....| 244
245 .....| 245
246 .....| 246
247 .....| 247
248 .....| 248
249 .....| 249
250 .....| 250
251 .....| 251
252 .....| 252
253 .....| 253
254 .....| 254
255 .....| 255
256 .....| 256
257 .....| 257
258 .....| 258
259 .....| 259
260 .....| 260
261 .....| 261
262 .....| 262
263 .....| 263
264 .....| 264
265 .....| 265
266 .....| 266
267 .....| 267
268 .....| 268
269 .....| 269
270 .....| 270
271 .....| 271
272 .....| 272
273 .....| 273
274 .....| 274
275 .....| 275
276 .....| 276
277 .....| 277
278 .....| 278
279 .....| 279
280 .....| 280
281 .....| 281
282 .....| 282
283 .....| 283
284 .....| 284
285 .....| 285
286 .....| 286
287 .....| 287
288 .....| 288
289 .....| 289
290 .....| 290
291 .....| 291
292 .....| 292
293 .....| 293
294 .....| 294
295 .....| 295
296 .....| 296
297 .....| 297
298 .....| 298
299 .....| 299
300 .....| 300
301 .....| 301
302 .....| 302
303 .....| 303
304 .....| 304
305 .....| 305
306 .....| 306
307 .....| 307
308 .....| 308
309 .....| 309
310 .....| 310
311 .....| 311
312 .....| 312
313 .....| 313
314 .....| 314
315 .....| 315
316 .....| 316
317 .....| 317
318 .....| 318
319 .....| 319
320 .....| 320
321 .....| 321
322 .....| 322
323 .....| 323
324 .....| 324
325 .....| 325
326 .....| 326
327 .....| 327
328 .....| 328
329 .....| 329
330 .....| 330
331 .....| 331
332 .....| 332
333 .....| 333
334 .....| 334
335 .....| 335
336 .....| 336
337 .....| 337
338 .....| 338
339 .....| 339
340 .....| 340
341 .....| 341
342 .....| 342
343 .....| 343
344 .....| 344
345 .....| 345
346 .....| 346
347 .....| 347
348 .....| 348
349 .....| 349
350 .....| 350
351 .....| 351
352 .....| 352
353 .....| 353
354 .....| 354
355 .....| 355
356 .....| 356
357 .....| 357
358 .....| 358
359 .....| 359
360 .....| 360
361 .....| 361
362 .....| 362
363 .....| 363
364 .....| 364
365 .....| 365
366 .....| 366
367 .....| 367
368 .....| 368
369 .....| 369
370 .....| 370
371 .....| 371
372 .....| 372
373 .....| 373
374 .....| 374
375 .....| 375
376 .....| 376
377 .....| 377
378 .....| 378
379 .....| 379
380 .....| 380
381 .....| 381
382 .....| 382
383 .....| 383
384 .....| 384
385 .....| 385
386 .....| 386
387 .....| 387
388 .....| 388
389 .....| 389
390 .....| 390
391 .....| 391
392 .....| 392
393 .....| 393
394 .....| 394
395 .....| 395
396 .....| 396
397 .....| 397
398 .....| 398
399 .....| 399
400 .....| 400
401 .....| 401
402 .....| 402
403 .....| 403
404 .....| 404
405 .....| 405
406 .....| 406
407 .....| 407
408 .....| 408
409 .....| 409
410 .....| 410
411 .....| 411
412 .....| 412
413 .....| 413
414 .....| 414
415 .....| 415
416 .....| 416
417 .....| 417
418 .....| 418
419 .....| 419
420 .....| 420
421 .....| 421
422 .....| 422
423 .....| 423
424 .....| 424
425 .....| 425
426 .....| 426
427 .....| 427
428 .....| 428
429 .....| 429
430 .....| 430
431 .....| 431
432 .....| 432
433 .....| 433
434 .....| 434
435 .....| 435
436 .....| 436
437 .....| 437
438 .....| 438
439 .....| 439
440 .....| 440
441 .....| 441
442 .....| 442
443 .....| 443
444 .....| 444
445 .....| 445
4
```

Observations	Number of observations	71
1	1	1
2	1	1
3	1	1
4	1	1
5	1	1
6	1	1
7	1	1
8	1	1
9	1	1
10	1	1
11	1	1
12	1	1
13	1	1
14	1	1
15	1	1
16	1	1
17	1	1
18	1	1
19	1	1
20	1	1
21	1	1
22	1	1
23	1	1
24	1	1
25	1	1
26	1	1
27	1	1
28	1	1
29	1	1
30	1	1
31	1	1
32	1	1
33	1	1
34	1	1
35	1	1
36	1	1
37	1	1
38	1	1
39	1	1
40	1	1
41	1	1
42	1	1
43	1	1
44	1	1
45	1	1
46	1	1
47	1	1
48	1	1
49	1	1
50	1	1
51	1	1
52	1	1
53	1	1
54	1	1
55	1	1
56	1	1
57	1	1
58	1	1
59	1	1
60	1	1
61	1	1
62	1	1
63	1	1
64	1	1
65	1	1
66	1	1
67	1	1
68	1	1
69	1	1
70	1	1
71	1	1
72	1	1
73	1	1
74	1	1
75	1	1
76	1	1
77	1	1
78	1	1
79	1	1
80	1	1
81	1	1
82	1	1
83	1	1
84	1	1
85	1	1
86	1	1
87	1	1
88	1	1
89	1	1
90	1	1
91	1	1
92	1	1
93	1	1
94	1	1
95	1	1
96	1	1
97	1	1
98	1	1
99	1	1
100	1	1
101	1	1
102	1	1
103	1	1
104	1	1
105	1	1
106	1	1
107	1	1
108	1	1
109	1	1
110	1	1
111	1	1
112	1	1
113	1	1
114	1	1
115	1	1
116	1	1
117	1	1
118	1	1
119	1	1
120	1	1
121	1	1
122	1	1
123	1	1
124	1	1
125	1	1
126	1	1
127	1	1
128	1	1
129	1	1
130	1	1
131	1	1
132	1	1
133	1	1
134	1	1
135	1	1
136	1	1
137	1	1
138	1	1
139	1	1
140	1	1
141	1	1
142	1	1
143	1	1
144	1	1
145	1	1
146	1	1
147	1	1
148	1	1
149	1	1
150	1	1
151	1	1
152	1	1
153	1	1
154	1	1
155	1	1
156	1	1
157	1	1
158	1	1
159	1	1
160	1	1
161	1	1
162	1	1
163	1	1
164	1	1
165	1	1
166	1	1
16		

[illegible]





a

• • •

2 AspargTrpLeuTrp..... 6  
 111 111 111  
 40676 GACCGCTGGAGCTGGCAGTAGATGGGCGCTGGGATATAGTTAGAGATGGG 40622

```

/organism="Mus musculus"
/strain="B6/CHAPLJ (C57 Black 6 x CBA)"
/db_xref="taxon:10090"
/clone="G20"
/clone_1fb="catalog #935302, Stratagene Cloning Systems,
La Jolla, CA"
/sex="female"
/tissue_type="liver"
<!. 84
85. 138
85. 2400
5'UTR
sig_peptide
CDS

```

```

/coodon_start=1
/function="binds and transports polymeric immunoglobulin"
/product="polymeric immunoglobulin receptor"
/protein_id="AA67440.1"
/db_xref="GI:458432"
/translation="MRTIYLTLLVYFSGYSTKSPFGQPEVSLSEGSISITCYPPD
ITSVNRKTRKCRGASGMOTLLIISNGYSKYSGRALINPENNFTYINIQILO
DQSGYSKGLGTLNRGSLSPVSLSEVSGVELPBDHYIYNDIGRNTIECFKRENP

```



BASE COUNT	ORIGIN
3' UTR	861 a 796 c 784 g 654 t
misc_feature	2401..3095
misc_feature	2089..2397
misc_feature	2020..2068
misc_feature	139..2019
mat_peptide	139..2397

```

alignment_scores:
    Quality: 198.00    Length: 168
    Ratio: 1.980      Gaps: 5
    Percent Similarity: 59.524    Percent Identity: 31.548

```

alignment\_block:  
US-09-135-238B-2 x MMU06431

Align seg 1/1 to: MMU06431 from: 1 to: 3095

[illegible]

1111  
MCCC S64  
E67

seq name: gb om:AF091137

```
seq_documentation_block;
```

DEFINITION	Trichosurus vulpecula	polymeric-immunoglobulin receptor precursor
LOCUS	AF091137	2848 bp
FEATURES	MRNA	MAM
DATE	01-JUL-1999	

(PIGR) mRNA, complete cds.

VERSION AF091137.1 GI:5305512  
KEYWORDS

SOURCE	Trichosurus vulpecula.
ORGANISM	Trichosurus vulpecula

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Metatheria; Diprotodontia; Phalangeridae; Trichosurus.

REFERENCE 1 (bases 1 to 2848)

TITLE Cloning and characterization of pIgR and J chain of the marsupial,

*Trichosurus vulpecula*  
Трошчатый барбас

REFERENCE 2 (bases 1 to 2848)  
AUTHORS Adamski, F. M. and Demmer, J.

**TITLE** Direct Submission  
**JOURNAL** Submitted (09-SEP-1998) Dairy Science, AgResearch Ruakura, East

screen, Hamilton, New Zealand  
Location/Qualifiers

SOURCE

**CDS**

```

/predict="polymeric-immunoglobulin receptor precursor"
/product_id="Abda1688.1"
/db_xref="GI:5305513"
translation="MAFFLACLALLAPVYSKMSPIFGKQVTCVGGSGVSIGCFYSS
SVNRGRVYFCLNLRSCETIVSSNGFVSERSGAKITNPGNSSLLIDISLENE
DIGLYKCGELGSNGLSPDITLLEVGEGPGLNPNTLEYTEGKVSINCPEDQND
RFLFKKDGESCAVLIDSOQVDPVPTGARLISGTSRFVYVTSIQIRBOVMYV
CGVGDSDSTGIQKNVDYDLKLEPEPLLVELGSSVYLNCAGDSIVASVPTFLQMAK
EEDVATINSKGFNNMATHRIILFSHTTEGSSINTQVRREGLSVFHCAGDEQGPS
EESGIRALQLEVESEETVPSKPLVYGGSGGVITLCHDRKNNTKLYKNGSSSH
CKVLDSLGMTDESEGRVALMDPEENGIFVYLNLDPDQAGKNTLCKNGENRRKS
VYIELNDQCPILPKPTVTLQDQSLTICSHCPKCYFVETKWCNMSOGCEITLPOE
BSSQAPFQDNQNSRNLSTLNSGTBDHGKAGWCNCKNGYMGVETLVSVEEYSG
NALQPTNVLNLEDVEPRKGRGKILVPTLQSGSTPHSGSSVLSLTVPLATLYAGA
VALGILKRRNRKESDRVSVGSTRIDLSBELDENPGRGAMENADSVQETILGEGDE
IITQENEEELQQRKSSKEEAMATLALDLSNNANNEILPDDPSNA"

```

```
alignment_scores:
  Quality: 192.00
  Ratio: 1.684
  Percent Similarity: 55.882
  Length: 204
  Gaps: 8
  Percent Identity: 30.392
```

```
alignment_block:
```

Align seg 1/1 to: AF091137 from: 1 to: 2848

```

17 LeuValIleLeuProGluVal..... 23
   ||| ..... |||
173 CTTCGCTCTCTCTCTGTTCTCTATGAGACTCCCATATTGGCCCCA 222
   ||| ..... |||

24 LysValGluGlyGluLeuGlyGlySerValThrIleLysCysProLeuP 40
   ..... |||
223 GCAAGTGACACGAGAGTGGAGAGGAGATCTGTCCATCCAGAGCTTAC 272
   ..... |||

```

















34 hrleLysCysProLeuProGluMetHisValArgIleTyrLeuCysArg 50  
 246 CCATCAAGTCCGACCTCTCTGAAATGATGATGAGGATATATCTGCGCG 295  
 51 GluMetAlaGlySerGlyThrCysGlyThrValSerThrThrAsnPh 67  
 296 GAGATGCTGATCTGGAACATCTGTACCTGTTATCCACCAACTT 345  
 67 eileLysAlaGlyThrGlyGlyValThrLeuLysGlnThrProArgL 84  
 346 CATCAAGGAGAAATCAAGGCGGAGTACTCTGAAGCAATACCCAGCA 395  
 84 yAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerGly 100  
 396 AGAATCTGTTCTAGTGGAGGTAAACACAGCTGACAGAAAGTGACAGCGGA 445  
 101 ValThrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnLy 117  
 446 GTCTATGCTGCGGAGCGGATGAAACACAGACCGGGGAAAGACCCAGAA 495  
 17 svalThrLeuAsnValHisSerGluThrGluProSerTrpGluGlnP 134  
 496 AGTCACCTGAATGTCCACAGTGAATACGAGCATCATGGGAGAGCAGC 545  
 134 roMetProGluThrProLysTrpPheHisLeuProTyrLeuPheGlnMet 150  
 546 CAATGCTGAGACTCAAAATGGTTTCATCTGCTCCCTATTGTTCCAGATG 595  
 151 ProAlaTyrAlaSerSerSerLysPheValThrArgValThrThrProAl 167  
 596 CCTGCATATGCCAGTCTCTCCAAATTCGTAAACAGAGTTACACACAGC 645  
 167 aGlnArgGlyLysValProProValHisHisSerSerProThrThrGlnI 184  
 646 TCAAGGGGAGAGTCCCTCCAGTTCACCATCTCCCTCCCTCCCTCCCTCA 695  
 184 lethrHisArgProArgValSerArgAlaSerSerValAlaGlyAspLys 200  
 696 TCACCCACCCCTCCAGTGTCCAGAGCATCTTCAGTAGCAGGTGACAAG 745  
 201 ProArgThrPheLeuProSerThrThrAlaSerLysIleSerAlaLeuG 217  
 746 CCCCAGAACCTTCTGCTGCTCCACTACAGCTCAAAATCTCAGTCTGGA 795  
 217 uGlyLeuLeuLysProGlnThrProSerTyrAsnHisHisThrArgLeuH 234  
 796 GGGGCTCTCAAGCCGACAGCCAGCCAGTACAAACCCACACACAGGCTGC 845  
 234 lsArgGlnArgAlaLeuAspTyrGlySerGlnSerGlyArgGluGln 250  
 846 ACAGCCAGAGACACTGGAGTATGCTACAGTCTGGGAGGAGAGGCCAA 895  
 251 GlyPheHisIleLeuLeuProThrLeuGlyLeuPheLeuAlaLe 267  
 896 GGATTTTCATCTGATCCGACCATCTGCGGCTTTTCTCTGCTGCACT 945  
 267 uLeuGlyLeuValValAlaArgAlaValAlaArgAlaLeuSerA 284  
 946 TCTGGGCTGCTGCTGTAAGGCGGCTTCAAGAGGAGGAAACCTCTCCA 995  
 284 rArgAlaArgArgLeuAlaValArgMetArgAlaLeuGluSerSerGln 300  
 996 GCGGGGCGCCGACCTGGCGGTGAGGATGCGGCTGAGAGGCTCCAG 1045  
 301 ArgProArgGlySerProArgProArgSerGlnAsnAsnIleTyrSerAl 317  
 1046 AGGCCCCGCGGCTGCGGCGGAGCGGCTCCCAAAACACATCTACAGCG 1095  
 317 acYsProArgArgAlaArgGlyAlaAspAlaAlaGlyThrGlyGlnAlap 334  
 1096 CTGCCCCGCGGCGCTGCTGGAGCGGAGCGCTCAGAGGACAGGGAGGCC 1145

334 roValProGlyProGlyAlaProLeuProProAlaProLeuGlnValSer 350  
 1146 CCGTTCCCGGCGGAGCGGTTGCCCCCGCCGCGCGTGCAGGTGCT 1195  
 351 GluSerProTriPLeuHisAlaProSerLeuLysThrSerCysGlyTrpVa 367  
 1196 GAATCTCCCTGGCTCATGCCCATCTCTGAAGACCACTGCTGAATACGT 1245  
 367 lSerLeuTyrHisGlnProAlaAlaMetMetGluAspSerAspSerAspA 384  
 1246 GAGCTCTACCCAGCGCTGCGCCCATCATGAGAGCAGATTCAGATG 1295  
 384 spTyrLeuAsnValProAla 390  
 1296 ACTACATCAATGTTCTGCTGCC 1315

seq\_name: N\_Geneseq\_36:X28179

seq\_documentation\_block:

ID X28179 standard; CDNA; 1047 BP.

AC X28179;

DT 16-JUN-1999 (first entry)

DE Human PIGRL-1 coding sequence fragment.

KW PIGRL-1; human; autoimmune disease; hyper-Igm Immunodeficiency; HIM;

KW X-linked Severe Combined Immunodeficiency; XSCID; IgA deficiency;

OS Homo sapiens.

OS Homo sapiens.

PN EP-905238-A2.

PD 31-MAR-1999.

PF 14-AUG-1998; 306487.

PR 30-OCT-1997; US-961564.

PR 25-AUG-1997; US-056935.

PA (SMK ) SMITHKLINE BEECHAM CORP.

PI Sweet RW, Truneh A, Wu S;

DR WPI: 99-192666/17.

DR P-PSDB; Y05002.

PT New polypeptides encoding PIGRL-1 useful for treating diseases such

as X-linked Severe Combined Immunodeficiency

PS Disclosure: Page 8; 26pp; English.

CC This sequence encodes the human PIGRL-1 protein of the invention.

CC Autoimmune diseases involving altered expression or activity of PIGRL-1

CC may include hyper-Igm Immunodeficiency (HIM), X-linked Severe Combined

CC Immunodeficiency (XSCID) and IgA deficiency. These diseases can be

CC diagnosed or susceptibility to them predicted by: (1) determining whether

CC there is a mutation in the genomic copy of the gene encoding PIGRL-1; or

CC (2) measuring the amount of PIGRL-1 in a sample derived from the patient.

CC Patients deficient in PIGRL-1 can be treated by administering either the

CC PIGRL-1 DNA or its complement or an agonist of PIGRL-1 to the patient.

CC Patients with excessive expression or activity of PIGRL-1 can be treated

CC by administering an antagonist of PIGRL-1, an antisense nucleic acid

CC molecule which inhibits the expression of PIGRL-1 or administering

CC sufficient PIGRL-1 to compete with the endogenous activity. PIGRL-1 can

CC be used to identify its agonists by contacting a cell expressing PIGRL-1

CC with a candidate compound in the presence of a signal system and noting

CC the candidate as an agonist if a signal is produced. The same method can

CC be used to identify antagonists of PIGRL-1 but the presence of an

CC antagonist is indicated by a decrease in production of the signal.

CC Antibodies against PIGRL-1 may be used to isolate or identify clones

CC expressing PIGRL-1. Polynucleotides encoding PIGRL-1 may be used to

CC identify chromosomal mutations in the gene encoding PIGRL-1 in patients.

CC This information may then be correlated with the incidence of autoimmune

CC disease in those patients to identify whether the mutation causes the

CC disease.

Sequence 1047 BP; 255 A; 281 C; 276 G; 223 T;

alignment\_scores:

Quality: 1125.00

Ratio: 4.429

Percent Similarity: 92.364

Length: 275

Gaps: 12

Percent Identity: 90.182

alignment\_block:

US-09-135-238B-2 x X28179





Align seg 1/1 to: X28179 from: 1 to: 1047

```
1 MetAspArgTrpLeuTrpProLeuTyPheLeuProValSerGlyAlaLe 17
|||||
150 ATGGACCTCTGGCTTTGGCCACCTTACTCTCCAGTATCAGGGCCCT 199
|||||
17 uArgLeuProGluValLysValGluGlyGluLeuGlyGlySerValT 34
|||||
200 GAGGATCTCCCAAGTAAGTAGAGGGGGAGCTGGCGGATCAGTTA 249
|||||
34 hrLeLysCysProLeuProGluMetHisValArgIleTyLeuCysArg 50
|||||
250 CCATCAAGTCCCACTCTCTGAATGATGTGAGGATATATCTGTCCGG 299
|||||
51 GluMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsnPh 67
|||||
300 GAGATGCTGGATCTGGAACATGTGTGTGATCCACCAACCACTT 349
|||||
67 eileLysAlaGluTyLysGlyArgValThrLeuLysGlnTyProArgL 84
|||||
350 CATCAAGCGAGAATACAGGGCGGAGTTACTCTGAAGCAATACCCACG 399
|||||
84 ysAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerGly 100
|||||
400 AGAATCTGTCTCTAGTGGAGGTAAACAGCTGACAGAAAGTGACAGCG 449
|||||
101 ValTyAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGlnTy 117
|||||
450 GTCTATGCTCGCGGAGC.GGCATGAACACAGACCGGGGAAAGACCCAG 498
|||||
117 svalThrLeuAsnValHisSerGluTyGluProSerTrpGluGlnP 134
|||||
499 AGTCACCTCTGAATCTCCACAGTAGTAATACGAGCCATCATGGGAAG 548
|||||
134 roMetProGluThrProLysTrpPheHisLeuProTyLeuPheGlnMet 150
|||||
549 CAATGCTCTGAGATCCCAAAATGGTTCTATCTGCCCTATTGTTCAGAT 598
|||||
151 ProAlaTyAlaSerSerLysPheValThr...ArgValThrThrPro 166
|||||
599 CCTGCATATCGCGTTCTTCCACATTCGTAAACCGCAGAGTTACCAACA 648
|||||
167 Ala.GlnArgGlyLysValProProValHisHisSerSerProThrThr 183
|||||
649 GCTTCAAGGGGCAAGGTCTCCAGTTCACCACTCTCCCGCCACCC 698
|||||
183 lnlle.ThrHisArgPro.ArgValSerArgAlaSerSerValAlaGly 199
|||||
699 AAATTCACCCACCGCCCTTCGATGTCAGAGCATCTTCAGTAGCAGGTG 748
|||||
199 spLysProArg.ThrPheLeuProSerThrThrAlaSerLysLysSerAl 215
|||||
749 ACAAGCCCGAAACTTCTCTGCCATCCACTACAGCCTCAAAATCTCAGC 798
|||||
215 aLeuGluGlyLeuLeuLys...ProGlnThrProSerTyThrAsnHisH 231
|||||
799 TCTGGAAGGCTGTGTTCAAGCCCGCAAGAGCGCCAGCTACAACAN.CACA 847
|||||
231 hrArgLeuHisArgGlnArgAlaLeuAsp...TyLysSerGlnSerG 246
|||||
848 CCAGGCTGCACAGCAGAGAGCACTGGATACTTATGGNTACAGTCTGG 897
|||||
246 yArgGluGly.....GlnGlyPhe...HisIleLeuIlePro.ThrIle 259
|||||
898 G...GAGGGGAANGNCCAAGGATTTTTCATCTCTGATTCCTCCGCGCA 944
|||||
260 LeuGlyLeuPhe 263
|||||
945 NTTGGGCGCTTT 956
```

seq\_name: N\_Geneseq\_36:T31290

seq\_documentation\_block:

```
ID T31290 standard; CDNA; 3095 BP.
AC T31290;
DT 24-FEB-1997 (first entry)
DE Mouse poly-immunoglobulin receptor, cDNA.
KW Mouse; immunoglobulin; receptor; protection protein; mutants;
KW heavy chain; antigen binding domain; protection; pathogen;
KW mucosal; environment; gastrointestinal; passive; immunisation;
KW Guy's 13 antibody; prevention; dental caries; Streptococcus;
KW poly; sorbinus; murine; ss.
OS Mus musculus.
FH Key Location/Qualifiers
FT cds 85..2400
FT /tag= a
FT WO9621012-A1.
PD 11-JUL-1996.
PF 27-DEC-1995; U16889.
PR 30-DEC-1994; US-367395.
PR 04-MAY-1995; US-434000.
PA (PLAN-) PLANT BIOTECHNOLOGY INC.
PA (UNME-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.
PA (PLAN-) PLANET BIOTECHNOLOGY INC.
PI Hiatt AC, Lehner T, Ma JKC.
DR WPI: 96-333987/33.
DR P-PSDB; W03180.
PT Immunoglobulin and protection protein complex and its prodn. in
PT plants - useful for passive immunisation against mucosal antigens,
PT esp. against S. mutans and S. sorbinus to prevent dental caries
PS Disclosure: pages 117-121; 152pp; English.
CC The present sequence encodes the mouse poly-immunoglobulin (Ig)
CC receptor, a portion of which corresp. to residues 1-627, pref.
CC 1-606, or esp. residues 13-45, 1-120, 110-230, 210-340, 320-450,
CC 440-550, 550-606 or 550-627 comprises a protection protein (PP).
CC The Ig of the invention comprises a PP as above in association with
CC an Ig derived heavy chain, having at least a portion of an antigen
CC (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.
CC gastrointestinal, environments, therefore enhancing its
CC effectiveness in passively immunising animals against mucosal
CC pathogens. The Ag binding domain is specifically derived from the
CC Guy's 13 antibody, and the Ig can be used to prevent dental caries
CC by binding, e.g. Streptococcus mutans serotypes C, e and f, or
CC S. sorbinus serotypes d and g.
SQ Sequence 3095 BP; 861 A; 796 C; 784 G; 654 T;
```

alignment\_scores:  
Quality: 198.00 Length: 168  
Ratio: 1.980 Gaps: 5  
Percent Similarity: 59.524 Percent Identity: 31.548

alignment\_block:  
US-09-135-238B-2 x T31290 ..

Align seg 1/1 to: T31290 from: 1 to: 3095

```
10 PheLeuProValSerGlyAlaLeuArgIleLeuProGluValLysValG 26
|||||
121 TTTTCAGGGGTCTCCACAAAAGCCCATATTTGGTCCCGCAGAGGTGAG 170
|||||
26 uGlyLeuLeuGlyGlySerValThrIleLysCysProLeuProGlu... 41
|||||
171 TAGTATAGAAGCGACTCTGTTCATCAGCTGCTACTACCCAGCACCT 220
|||||
42 .....MetHisValArgIleTyLeuCysArgGluMetAlaGlySer 55
|||||
221 CTGTCAACCGGCACACCGGAAATACTGGTCCGCAAGGAGCC...AGC 267
|||||
56 GlyThrCysGlyThrValValSerThrThrAsnPheIleLysAlaGluTy 72
|||||
268 GGCATGTGCACACGCTCATCTCTCAATGGCTACCTCTCCAAGGAGTA 317
|||||
72 rLysGlyArgValThrLeuLysGlnTyProArgLysAsnLeuPheLeuV 89
|||||
```







QY 940 ggcgtgagatgcgcgcctcggagagctccagagagcccgccgctgcgcgcgacccg 999  
 Db 411 ggcgtgagatgcgcgcctcggagagctccagagagcccgccgctgcgcgcgacccg 352  
 QY 1000 ctcccaaaacacatctacagcgcctgcgcgcgcgcgcctc-tgagagacacactgcagg 1058  
 Db 351 ctcccaaaacacatctacagcgcctgcgcgcgcgcgcctc-tgagagacacactgcagg 292  
 QY 1059 cacaggggagggcccgctcccgcccgagcgcgcctgcccccccgcccgctgcagggt 1118  
 Db 291 cacaggggagggcccgctcccgcccgagcgcgcctgcccccccgcccgctgcagggt 232  
 QY 1119 gtcgaatctccctggtccatgcccatctctgaagaccagctgtgaatacgtgagcct 1178  
 Db 231 gtttgaatctccctggtccatgcccatctctgaagaccagctgtgaatacgtgagcct 172  
 QY 1179 ctaccacagcctcccgccatgagggagcagtgattcagatgactacatcaatgtcc 1238  
 Db 171 ctaccacagcctcccgccatgagggagcagtgattcagatgactacatcaatgtcc 112  
 QY 1239 tgcctgacaactcccgagctatcccccaacccagcgcgcctggtgcccagagtc 1298  
 Db 111 tgcctgacaactcccgagctatcccccaacccagcgcgcctggtgcccagagtc 52  
 QY 1299 tcactatctgctgagtgatcaaatcactgttcctatgtgtcttcaga 1343  
 Db 51 tcactatctgctgagtgatcaaatcactgttcctatgtgtcttcaga 7

RESULT 2  
 AI084226/c  
 LOCUS AI084226 578 bp mRNA EST 24-SEP-1998  
 DEFINITION Oy72g09.x1 NCI\_CGAP\_CLL1 Homo sapiens cDNA clone IMAGE:1671424 3', mRNA sequence.  
 ACCESSION AI084226  
 VERSION AI084226.1 GI:3422649  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 578)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On Jun 5, 1998 this sequence version replaced gi:3189667.  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever, M.D., Louis M. Staudt, M.D., Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html  
 Insert Length: 940 Std Error: 0.00  
 Seq primer: -40ml3 fwd. ET from Amersham  
 High quality sequence stop: 438.  
 Location/Qualifiers  
 1..578  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1671424"  
 /clone\_lib="NCI-CGAP\_CLL1"  
 /tissue\_type="B-cell, chronic lymphocytic leukemia"  
 /lab\_host="DH10B"  
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not 1; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGAGCGCGCCATGCTTTTTTTTTTTTTTTTTT

# FEATURES

source  
 1..578  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1671424"  
 /clone\_lib="NCI-CGAP\_CLL1"  
 /tissue\_type="B-cell, chronic lymphocytic leukemia"  
 /lab\_host="DH10B"  
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not 1; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGAGCGCGCCATGCTTTTTTTTTTTTTTTTTT

T 3'; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized, and was constructed by Bento Soares and M.Fatima Bonaldo.  
 BASE COUNT 160 a 123 c 156 g 139 t  
 ORIGIN

Query Match 27.0%; Score 516.4; DB 35; Length 578;  
 Best Local Similarity 96.6%; Pred. No. 4.7e-131;  
 Matches 560; Conservative 0; Mismatches 16; Indels 4; Gaps 3;  
 QY 1299 tcactatctgctgagtgatcccaataaccctgcttcctgcttcagagccctcatcactcc 1358  
 Db 578 TCATCTATCTGCTGATGTCCATACCTGCTTCATGTGTCTCAGAGCCCTCATCACTCC 519  
 QY 1359 catgcccactctcgactccatccatccatctatctggtccctgagcagctgctgcctcc 1418  
 Db 518 CATGCCCATCTCGACTCCCATCCCATCTATCTGT-GCCCTGAGCATGGCTCTGCCCCC 460  
 QY 1419 aggtcgctttgcaacacctggcagccctctgattgacagagtgatgagtgatgag 1478  
 Db 459 AGGTGCTCTTGCACACCTTGGCAGCCCTCTGTAGTTGACAGGTAAGCTGTAGGCATGTAG 400  
 QY 1479 agcaattgtcccaatgccactgtcttcccaagccgtcgaaacagactgtggatttg 1538  
 Db 399 AGCAANTGTCCCATGCCACTTGTCTTCCAAAGCCGTGAAACAGACTGTGGGATTTG 340  
 QY 1539 cagagtgctttccatgctcttggaccacaggggtgtgtgtgtgtgtgtgtgtgtgtgtgt 1598  
 Db 339 CAGAGTGTCTTCTCCATGCTTTGACACACAGGT-TGTGTGTGCCAGGCTCTAGATCAC 281  
 QY 1599 atggcatcaggtggggcagagggcagcattgtctctggggcctctcccaagggttggg 1658  
 Db 280 ATGGCATCAGGCTGGGGCAGAGGATAGCTATTGTCTGGGCATCCTTCCAGGGTTGGG 221  
 QY 1659 tcttacacaaatagaaggctctgtctgtgattgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1718  
 Db 220 TCTTACACAAATAGAGGCTCTTGTCTGTAGTTATGTGACATGCTCAGCCCCATGACT 161  
 QY 1719 aagcagggtgtgtgtat--aaacactctggaaacgcctttgcccctgacataatgttag 1776  
 Db 160 AAGCAGGGGTCTGGTATATAAAACACTCTGGAACGCTTTGCCCTGATCCAAATGTTAG 101  
 QY 1777 caattgttagtgaaacgtctacttacttacttacttacttacttacttacttacttacttactt 1836  
 Db 100 CACTGTGTAGTGAACGCTTACTTATCTCAAGTTCTATGCTAAAGGCAATTTATCTTGATG 41  
 QY 1837 tgatgataaaccaactattattagcaagatatgcataata 1876  
 Db 40 TGATGATAAACCAACTATTATTAGCAAAAAAATAAAAAA 1

RESULT 3  
 AW452045/c  
 LOCUS AW452045 524 bp mRNA EST 17-FEB-2000  
 DEFINITION UI-H-B13-alm-h-05-0-UI.s1 NCI\_CGAP\_Sub5 Homo sapiens cDNA clone IMAGE:2737400 3', mRNA sequence.  
 ACCESSION AW452045  
 VERSION AW452045.1 GI:6992821  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 524)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On Jul 7, 1999 this sequence version replaced gi:5405901.  
 Contact: Robert Strausberg, Ph.D.



LOCUS AW402953 312 bp mRNA 16-FEB-2000  
DEFINITION UI-HF-BK0-abd-c-06-0-UI.r1 NIH\_MGC\_36 Homo sapiens cDNA clone  
IMAGE:3055859 5', mRNA sequence.  
ACCESSION AW402953  
VERSION AW402953.1 GI:6921743  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 312)  
AUTHORS NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT On May 18, 1998 this sequence version replaced gi:3138550.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1350  
Email: Robert.Strausberg@nih.gov  
Eco RI site shown at the beginning of the sequence.  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
CDNA Library Preparation: M.B. Soares Lab  
CDNA Library Arrayed by: M.B. Soares Lab  
DNA Sequencing by: M.B. Soares Lab  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/dbp/image/image.html](http://www.bio.llnl.gov/dbp/image/image.html)  
Seq primer: M3 Forward

FEATURES  
source  
1..312  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3055859"  
/clone\_lib="NIH\_MGC\_36"  
/tissue\_type="lymph"  
/cell\_type="germinal center B cells"  
/lab\_host="MGC85"  
/note="Vector: pT7T3-Pac; site\_1: NotI; site\_2: Eco RI;  
Constructed from size fractionated cytoplasmic mRNA  
(0.5-1.5kb). Directionally cloned. Cells provided by Louis  
M. Staudt, Ph.D. Library preparation by Maria de Fatima  
Bonaldo, Ph.D. and M. Bento Soares, Ph.D."  
BASE COUNT 97 a 74 c 87 g 54 t

alignment\_scores  
Quality: 544.00 Length: 101  
Ratio: 5.386 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-09-135-238B-2 x AW402953 ..

Align seg 1/1 to: AW402953 from: 1 to: 312  
41 GluMetHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyth 57  
10 GAAATGATGTGAGGATATATCTGTCCCGGAGATGGCTGGATCTGGAAC 59  
57 rCysGlyThrValValSerThrThrAsnPheIleLysAlaGluTyrLys 74  
60 ATGTGGTACCGTGGTATCCACACCACTCATCAAGGAGATACAAAG 109  
74 lyArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGlu 90  
110 GCCGAGTTACTCTGAACAAATATCCACGCAAGATCTGTCTCTAGTGGAG 159  
91 ValThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaG 107  
160 GTAACACAGCTGACAGAAAGTACACGGAGTCTATGCTCGCGGCGGG 209  
107 yMetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnValHis 124

|||||  
210 CATGACACAGACCGGGAAAGACCCAGAAAGTACCCCTGATGTCCACA 259  
124 erGluTyrGluProSerTrpGluGluInProMetProGluThrProLys 140  
260 GTGAATAGAGCCATCATGTGGGAAGACGACCAATGCCTGAGACTCCAAA 309  
141 Trp 141  
310 TGG 312  
seq\_name: gb\_est45:AW655210  
seq\_documentation\_block:  
LOCUS AW655210 518 bp mRNA EST 05-APR-2000  
DEFINITION 105863 MARC 1B0V Bos taurus cDNA 5', mRNA sequence.  
ACCESSION AW655210  
VERSION AW655210.1 GI:7421036  
KEYWORDS EST.  
SOURCE Bos taurus.  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
REFERENCE 1 (bases 1 to 518)  
AUTHORS Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,  
Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A. and  
Keele,J.W.  
TITLE Design and use of four pooled tissue normalized cDNA libraries for  
EST discovery in cattle  
JOURNAL Unpublished (2000)  
COMMENT On Mar 8, 1999 this sequence version replaced gi:4388119.  
Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@email.marc.usda.gov  
Single pass sequencing. Bases called and alt\_trimmed with phred  
v0.980904.e. Vector identified by cross\_match with the -minscore 18  
and -mismatch 12 options.  
PCR Primers  
FORWARD: AGGAAACAGCTATGACCAT  
BACKWARD: GTTTTCCCGTCACGACG  
Plate: 86 row: K column: 10  
Seq primer: ATTTAGTGACACTAG.

FEATURES  
source  
1..518  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_lib="MARC 1B0V"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/note="Vector: pCMV SPORT6; site\_1: XbaI; site\_2: XhoI;  
Library made from pooled tissue from lymph node, ovary,  
fat, hypothalamus, and pituitary."  
BASE COUNT 124 a 133 c 157 g 104 t  
ORIGIN

alignment\_scores  
Quality: 522.00 Length: 136  
Ratio: 4.424 Gaps: 0  
Percent Similarity: 86.765 Percent Identity: 68.382  
alignment\_block:  
US-09-135-238B-2 x AW655210 ..  
Align seg 1/1 to: AW655210 from: 1 to: 518  
1 MetAspArgTrpLeuTrpProLeuTyrPheLeuProValSerGlyAlaLe 17  
110 ATGGACCTGTGGCTTGGCGCTTTACTTCTCTCAGTAGTCGAGGCC 159





APR

LOCUS AW402953 312 bp mRNA EST 16-FEB-2000  
DEFINITION UI-HF-BK0-abd-c-06-0-UI.r1 NIH\_MGC\_36 Homo sapiens cDNA clone IMAGE:3055859 5', mRNA sequence.  
ACCESSION AW402953  
VERSION AW402953.1 GI:69211743  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 312)  
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)  
JOURNAL On May 18, 1998 this sequence version replaced gi:3138550.  
COMMENT Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
ECO RI site shown at the beginning of the sequence.  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
CDNA Library Preparation: M.B. Soares Lab  
CDNA Library Arrayed by: M.B. Soares Lab  
DNA Sequencing by: M.B. Soares Lab  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.lnli.gov/bbrp/image/image.html  
Seq primer: M13 Forward.  
FEATURES  
source  
1..312  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3055859"  
/clone\_lib="NIH\_MGC\_36"  
/tissue\_type="lymph"  
/cell\_type="germinal center B cells"  
/cell\_line="MC85"  
/lab\_host="DH10B (LTR)"  
/note="vector: p7T3-Pac; Site\_1: NotI; Site\_2: Eco RI; Constructed from size fractionated cytoplasmic mRNA (0.5-1.5kb). Directionally cloned. Cells provided by Louis M. Staudt, Ph.D. Library preparation by Maria de Fatima Bonaldo, Ph.D. and M. Bento Soares, Ph.D."  
BASE COUNT 97 a 74 c 87 g 54 t  
ORIGIN

alignment\_scores:  
Quality: 544.00 Length: 101  
Ratio: 5.386 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-35-2388-2 x AW402953  
Align seg 1/1 to: AW402953 from: 1 to: 312  
41 GluMetHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThr 57  
10 GAATGTCATGTGAGGATATATCTGTCGGGAGATGCTGTGATCTGGAAC 59  
57 rCysGlyThrValValSerThrThrAspPheIleLysAlaGluTyrLysG 74  
60 ATGTGGTACCGTGGTATCATCCACCACTTCATCAAGCAGACATCAAGG 109  
74 lYArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGlu 90  
110 GCGGAGTTACTCTGAACAATACCCACCAAGAAATCTGTTCTCTAGTGAG 159  
91 ValThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaG 107  
160 GTAACACAGCTGACAGAAAGTGCACAGCGAGTCTATGCTCGGAGCGGG 209  
107 yMetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnValHis 124

210 CATGAACACAGACCGGGGAAACACCCAGAAAGTCACCTGATGTCCACA 259  
124 erGluTyrGluProSerTrpGluGluGlnProMetProGluThrProLys 140  
260 GTGAATACAGCCATCATGTGGGAAGAGACGCCAATGCTGAGACTCAAAA 309  
141 Trp 141  
310 TGG 312  
seq\_name: gb\_est45:AW655210  
seq\_documentation\_block:  
LOCUS AW655210 518 bp mRNA EST 05-APR-2000  
DEFINITION 105863 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.  
ACCESSION AW655210  
VERSION AW655210.1 GI:7421036  
KEYWORDS EST.  
SOURCE Bos taurus.  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.  
REFERENCE 1 (bases 1 to 518)  
AUTHORS Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A. and Keeler,J.W.  
TITLE Design and use of four pooled tissue normalized cDNA libraries for EST discovery in cattle  
JOURNAL Unpublished (2000)  
COMMENT On Mar 8, 1999 this sequence version replaced gi:4388119.  
Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@mail.marc.usda.gov  
Single pass sequencing. Bases called and alt\_trimmed with phred v0.980904.e. Vector identified by cross\_match with the -minscore 18 and -mismatch 12 options.  
PCR primers  
FORWARD: AGGAACACAGCTATGACCAT  
BACKWARD: GTTTCACGATCAGCAGC  
Plate: 86 row: K column: 10  
Seq primer: ATTTAGGTGACACTATAG.  
FEATURES  
Location/Qualifiers  
1..518  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_lib="MARC 1BOV"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/note="Vector: pCMV SPORT6; Site\_1: XbaI; Site\_2: XhoI; Library made from pooled tissue from lymph node, ovary, fat, hypothalamus, and pituitary."  
BASE COUNT 124 a 133 c 157 g 104 t  
ORIGIN

alignment\_scores:  
Quality: 522.00 Length: 136  
Ratio: 4.424 Gaps: 0  
Percent Similarity: 86.765 Percent Identity: 68.382  
alignment\_block:  
US-09-135-2388-2 x AW655210  
Align seg 1/1 to: AW655210 from: 1 to: 518  
1 MetAspArgTrpLeuTrpProLeuTyrPheLeuProValSerGlyAlaLe 17  
110 ATGGACGTGTGGCTTTGGCGCTTTACTTCTTCCAGTAGTCGGAGCCCC 159





282 GAACGGCACCATTCTGCTGAACATGCCCCAGCTGAGCAGGATGACTCCG 331  
 83 lyValtyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGln 99  
 || ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 332 GCGCTACAAGTGTGGCTGGCATCAATAGC...CGAGGCTGTCTTT 378  
 100 LysValThrLeuAsnVal 105  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 379 GATGTCAGCCTGGAGGTC 396

seq\_name: N\_Geneseq\_36:T30856

seq\_documentation\_block:  
 ID T30856 standard; cDNA; 2031 BP.  
 AC T30856;  
 DT 13-SEP-1996 (first entry)  
 DE Secretory component DNA fragment 1.  
 KW Secretory component; SC DNA fragment 1; crystallisation;  
 KW polymeric immunoglobulin receptor; pIgR; CHO; ss.  
 OS Chimeric Homo sapiens;  
 OS Chimeric synthetic.  
 FH Key Location/Qualifiers  
 FT misc\_feature 1..33  
 FT /tag= a  
 FT /note= "bases 1-33 derived from PCR with 5'  
 primer 1"  
 FT primer 1  
 FT 111..2015  
 FT /tag= b  
 FT /product= transcript from SC DNA fragment 1  
 FT complement (1997..2031)  
 FT /tag= c  
 FT /note= "from PCR with 3' primer"  
 FT WO9618734-A1.  
 PN 20-JUN-1996.  
 PD 06-DEC-1995; E04797.  
 PR 16-DEC-1994; EP-120019.  
 PA (CIBA ) CIBA GEIGY AG.  
 PI Asselbergs FAM, Gruetter M, Leist C, Van Oostrum J;  
 DR WPI; 96-300651/30.  
 DT Prod. of recombinant secretory protein capable of crystallisation -  
 pref. human poly-IgR, for use in screening and binding studies  
 PS Claim 5; Page 36-38; 50pp; English.  
 CC SC DNA fragment 1 (T30856) was obt'd. by PCR amplification of human  
 CC polymeric immunoglobulin receptor (pIgR) cDNA in vector pCB6 using a  
 CC 5' primer (T30859) hybridising to the cytomegalovirus promoter in  
 CC pCB6 and a 3' primer (T30861) that generates a stop codon at  
 CC position 1906 just upstream of the hydrophobic transmembrane segment  
 CC of the pIgR and also creates a downstream XbaI site. CHO SSF 3  
 CC cells transfected with a vector carrying the amplified DNA can be  
 CC used for prodn. of recombinant secretory component (SC). This can  
 CC be crystallised for receptor structure studies, used to stabilise  
 CC immunoglobulins, or used to screen (ant)agonists capable of  
 CC modulating mucosal immune responses.  
 SQ Sequence 2031 BP; 477 A; 541 C; 605 G; 408 T;

alignment\_scores:  
 Quality: 180.50 Length: 106  
 Ratio: 2.314 Gaps: 4  
 Percent Similarity: 73.585 Percent Identity: 38.679

alignment\_block:

US-09-135-238b-2\_copy\_18\_253 x T30856 ..

Align seq 1/1 to: T30856 from: 1 to: 2031

4 ProGluValLysValGluGlyGluLeuValThrIleIysCys 20  
 183 CCGAG...GAGGTGAATAGTGTGGAGGTAACAGTCACTGTCATCGTG 229  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 20 sProLeuPro.....GluMetHisValArgIleTyrLeuCysA 33  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 230 CTACTACCCACCCACTCTGTCAACCGCGACACCCCGAAGTACTGTGTC 279

33 rgGluMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsn 49  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 280 GCGAGGAGGTAGAGGTGGC...TGCATAACCCCTCATCTCTCGGAGGCG 326  
 50 PheIleLysAlaGluTyrLysGlyValGValThrLeuLysGlnTyrProAr 66  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 327 TACGTCCTCCAGCAATATGCGAGCAGGGCTAACCTCACCACACTCCCGGA 376  
 66 GlyAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerG 83  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 377 GAACGGCACATTCTGCTGTAACATTGCCAGCTGAGCCAGGATGACTCCG 426  
 83 lyValTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGln 99  
 || ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 427 GGCCTACAGTGTGGCTGGCATCANTAGC...CGAGGCTGTCTTT 473  
 100 LysValThrLeuAsnVal 105  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 474 GATGTCAGCCTGGAGGTC 491

seq\_name: N\_Geneseq\_36:T31288

seq\_documentation\_block:  
 ID T31288 standard; cDNA; 2919 BP.  
 AC T31288;  
 DT 24-FEB-1997 (first entry)  
 DE Human poly-immunoglobulin receptor, cDNA.  
 KW Human; immunoglobulin receptor; protection protein; mutants;  
 KW heavy chain; antigen binding domain; protection; pathogen;  
 KW mucosal; environment; gastrointestinal; passive; immunisation;  
 KW Guy's 13 antibody; prevention; dental caries; Streptococcus;  
 KW poly; sorbinus; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT cds 235..2475  
 FT /\*tag= a  
 FT WO9621012-A1.  
 PN 11-JUL-1996.  
 PD 27-DEC-1995; U16889.  
 PR 30-DEC-1994; US-367395.  
 PR 04-MAY-1995; US-434000.  
 PA (PLAN-) PLANT BIOTECHNOLOGY INC.  
 PA (UNME-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.  
 PA (PLAN-) PLANET BIOTECHNOLOGY INC.  
 PI Hiatt AC, Leher T, Ma JKC;  
 DR WPI; 96-333987/33.  
 DR P-PSDB; W03178.  
 DT Immunoglobulin and protection protein complex and its prodn. in  
 PT plants - useful for passive immunisation against mucosal antigens,  
 PT esp. against S. mutants and S. sorbinus to prevent dental caries  
 PT esp. against S. mutants and S. sorbinus to prevent dental caries  
 PS Disclosure: Pages 105-108; 152pp; English.  
 CC The present sequence encodes the human poly-immunoglobulin (Ig)  
 CC receptor, a portion of which corresp. to residues 1-627, pref.  
 CC 1-606, or esp. residues 13-45, 1-120, 110-230, 210-340, 320-450,  
 CC 440-550, 550-606 or 550-627 comprises a protection protein (PP).  
 CC The Ig of the invention comprises a PP as above in association with  
 CC an Ig derived heavy chain, having at least a portion of an antigen  
 CC (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.  
 CC gastrointestinal, environments, therefore enhancing its  
 CC effectiveness in passively immunising animals against mucosal  
 CC pathogens. The Ag binding domain is specifically derived from the  
 CC Guy's 13 antibody, and the Ig can be used to prevent dental caries  
 CC by binding, e.g. Streptococcus mutants serotypes c, e and f, or  
 CC S. sorbinus serotypes d and g.  
 SQ Sequence 2919 BP; 684 A; 806 C; 854 G; 575 T;

alignment\_scores:  
 Quality: 180.50 Length: 106  
 Ratio: 2.314 Gaps: 4  
 Percent Similarity: 73.585 Percent Identity: 38.679

## alignment\_block:

```
US-09-135-238B-2_COPY_18_253 x T31288 ..
Align seg 1/1 to: T31288 from: 1 to: 2919

4 ProGluValLysValGluGlyLeuGlySerValThrIleLysCys 20
||||| :|||:||||| :|||:|||||:||||| ||
253 CCGGAG...GAGGTGAATAGTGTGAAGTAACCTCAGTGCTCATCACGTG 299

20 sProLeuPro.....GluMetHisValArgIleTyrLeuCysA 33
| ||| ||| :|||:||||| :|||:||||| ||| |||
300 CTACTACCCACCACCTCTGTCAACGGCACACCGGAAGTACTGTGTCG 349

33 rgGluMetAlaGlySerGlyThrCysGlyThrValValSerThrAsn 49
|||: ||| :|||:||||| ||| :|||:|||||: ||| :|||:
350 GGCAGGGAGCTAGAGTGTC...TGCATAACCTCTATCTCTCGGAGGCG 396

50 PheIleLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProAr 66
:||||:|||||:|||||:|||||:|||||:|||||:|||||
397 TAGCTCTCCAGCAATATAGCAGGAGGCTAACCTCACCACTTCCCGGA 446

66 gLysAsnLeuPheLeuValGluValThrClnLeuThrGluSerAspSerg 83
:||||:|||||:|||||:|||||:|||||:|||||:|||||
447 GAACGGCACATTGTGGTGAACATTGCCAGCTGAGCCAGGATGACTCCG 496

83 lyValTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGln 99
|| ||| ||||| |||||:|||||: ||||| :|||
497 GCGCGTACAAGTGTGGCTGGGCATCATATAGC...CGAGGCGTGTCTTT 543

100 LysValThrLeuAsnVal 105
||||:|||||:|||||
544 GATGTCAGCGCTGGAGTGC 561
```

seq\_name: N\_Geneseq\_36:T31289

## seq\_documentation\_block:

```
ID T31289 standard; cdNA; 3630 BP.
AC T31289;
AT 24-FEB-1997 (first entry)
DE Bovine poly-immunoglobulin receptor, cdNA.
KW Bovine; immunoglobulin; receptor; protection protein; mutants;
KW heavy chain; antigen binding domain; protection; pathogen;
KW mucosal; environment; gastrointestinal; passive; immunisation;
KW Guy's 13 antibody; prevention; dental caries; Streptococcus;
KW poly; sorbinus; cow; ss.
OS Bos taurus.
FH key Location/Qualifiers
FT cds 152..2425
FT /*tag= a
PN WO9621012-A1.
PD 11-JUL-1996.
PF 27-DEC-1995; U15889.
PR 30-DEC-1994; US-367395.
PR 04-MAY-1995; US-434000.
PA (PLAN-) PLANT BIOTECHNOLOGY INC.
PA (UNNE-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.
PA (PLAN-) PLANET BIOTECHNOLOGY INC.
PI Hiatt AC, Lehner T, Ma JKC;
DR WPI; 96-333987/33.
DR P-PSDB; W03179.
PT Immunoglobulin and protection protein complex and its prodn. in
PT plants - useful for passive immunisation against mucosal antigens,
PT esp. against S. mutans and S. sorbinus to prevent dental caries
PS Disclosure; Pages 111-115; 152pp; English.
CC The present sequence encodes the bovine poly-immunoglobulin (Ig)
CC receptor, a portion of which corresp. to residues 1-627, pref.
CC 1-606, or esp. residues 13-45, 1-120, 110-230; 210-340, 320-450,
CC 440-550, 550-606 or 550-627 comprises a protection protein (PP).
CC The Ig of the invention comprises a PP as above in association with
CC an Ig derived heavy chain, having at least a portion of an antigen
CC (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.
CC gastrointestinal, environments, therefore enhancing its
CC effectiveness in passively immunising animals against mucosal
```

```
CC pathogens. The Ag binding domain is specifically derived from the
CC Guy's 13 antibody, and the Ig can be used to prevent dental caries
CC by binding, e.g. Streptococcus mutans serotypes c, e and f, or
CC S. sorbinus serotypes d and g.
SQ Sequence 3630 BP; 880 A; 1018 C; 1029 G; 703 T;
```

## alignment\_scores:

```
Quality: 179.50 Length: 109
Ratio: 2.272 Gaps: 4
Percent Similarity: 72.477 Percent Identity: 37.615
```

## alignment\_block:

US-09-135-238B-2\_COPY\_18\_253 x T31289 ..

Align seg 1/1 to: T31289 from: 1 to: 3630

```
4 ProGluValLysValGluGlyGluLeuGlyGlySerValThrIleLysCys 20
||||| :|||:||||| ||| :|||:|||||: ||| :|||:
224 CCGGAG...GAGGTGAGCAGCGTGGAAAGCGCGCTCAGTGCTCCATCAAGTG 270

20 sProLeuPro.....GluMetHisValArgIleTyrLeuCysA 33
| ||| ||| :|||:||||| ||| :|||:|||||: ||| :|||:
271 CTACTACCCGCCACCTCCGTCAACGGCACACGCGCAAGTACTGTGTCG 320

33 rgGluMetAlaGlySerGlyThrCysGlyThrValValSerThrThrAsn 49
||||: ||| :|||:||||| ||| :|||:|||||: ||| :|||:
321 GGCAGGAGGCC...CAGGGCGCGCTGCAGCACCTCATCTCTCGGAGGGC 367

50 PheIleLysAlaGluTyrLysGlyArgValThrLeuLysGlnTyrProAr 66
:||||:|||||:|||||:|||||:|||||:|||||:|||||
368 TAGCTCTCCGACGACTACGTGGCGCAGAGCAACCTCACCAACTTCCCGGA 417

66 gLysAsnLeuPheLeuValGluValThrGlnLeuThrGluSerAspSerg 83
:||||:|||||:|||||:|||||:|||||:|||||:|||||
418 GAGCGGCACGTTTGTGTGGTGACATCAGCCATCTCACCCATAAAGACTCAG 467

83 lyValTyrAlaCysGlyAlaGlyMetAsnThrAspArgGlyLysThrGln 99
|| ||| ||||| |||||:|||||: ||||| :|||
468 GCGCGTACAAGTGTGGCTGGGCATTAGCAGC...CGTGGCGCTTAAC TTC 514

100 LysValThrLeuAsnValHisSergGlu 108
||||:|||||:|||||
515 GATGTCAGCGCTGGAGGTCAGCCCAAGAT 541
```

seq\_name: N\_Geneseq\_36:V20383

## seq\_documentation\_block:

```
ID V20383 standard; cdNA to mRNA; 1911 BP.
AC V20383;
AT 26-JUN-1998 (first entry)
DE cdNA for human immunity related factor.
KW Lymph node; human; immunity related factor; research; treatment;
KW immune disease; infectious disease; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 98..1696
FT /*tag= a
FT sig_peptide 98..145
FT /*tag= b
FT mat_peptide 146..1693
FT /*tag= c
FT /product= immunity_related_factor
PN J10072495-A.
PD 17-MAR-1998.
PF 11-JUN-1997; 153218.
PR 13-JUN-1996; JP-152362.
PA (ASAH ) ASAH KASEI KOGYO KK.
DR WPI; 98-234766/21.
DR P-PSDB; W50033.
PT Immunity related factor - useful in the treatment of immune related
PT and infectious diseases
PS Claim 15; Pages 18-20; 21pp; Japanese.
```

CC The present sequence encodes a lymph node derived human immunity  
CC related factor, which can be used to research and treat immune and  
CC infectious diseases.  
SQ Sequence 1911 BP: 490 A; 541 C; 525 G; 355 T;

alignment\_scores:  
Quality: 175.50 Length: 306  
Ratio: 1.415 Gaps: 10  
Percent Similarity: 40.523 Percent Identity: 21.895

alignment\_block:

US-09-135-238B-2\_COPY\_18\_253 x V20383 ..

Align seg 1/1 to: V20383 from: 1 to: 1911

```
8 ValGluGlyGluLeuLeuGlySerValThrIleLysCysProLeuPro.. 23
|||||
305 GTGTGAGGGAGCGCTGGAGAGCTGTCCACCATCGTGCCTATATGCC 354
24 .....GluMethHisValArgIleTyrLeuCysArgGluMetalag 37
   :||| ||| ||| ||| |||
355 CTCATCTGTCAACAGGCACGACGAGAGTACTGTGCGCTGTGGGCC 404
37 LysSerGlyThrCysGlyThrValValSerThrThrAsnPheIleLysAla 53
   ||| ||| ||| ||| |||
405 CAAGATGGATCTGCCAGACCATGTGTCCACCAACCATATATCA 454
54 GluTyrLysGlyArgValThrLeuLysGlnTyrProArgLysAsnLeuPh 70
   :||| ||| ||| ||| |||
455 CGCTATCTGACCTGTGGCCCTACACACTTCCACAGAGAGCTGTT 504
70 eLeuValGluValThrGlnLeuThrGluSerAspSerGlyValTyrAla 87
   ||| ||| ||| ||| |||
505 TGTGTGAGGCTGTCCCAACTGTCCCGGATGACATCGATCTACTCT 554
87 YsGlyAlaGlyMetAsnThrAspArg..... 95
   ||| ||| ||| ||| |||
555 GCGGATTTGGAGTGAACACACACTGTGTTCTTAAGCATGAATGACC 604
95 ..... 95
605 ATCTCTGAGGTCCCGCCAGCACCTCCACAGCCACTCCAGCTGTGG 654
95 ..... 95
655 GGAGCTCACCATGAGATCTTATGGACAGCGTCTCCAGTGGCCACAGAT 704
96 .....GlyLysThrGln..... 99
   ||| ||| ||| |||
705 GGACCCCGAGGAACCCAGACCTTAGGACAGGGGACAGCATGGGACACA 754
100 .....LysValThrLeuAsnValHisSerGln 108
   ||| ||| ||| ||| |||
755 GTGTCTTCATCTCAGACACCGACGAGACTACACTTCAGCTGAGGAG 804
108 uTyrGluPro.....SerTrp 114
   ||| ||| ||| |||
805 ACGAACCCCGAGGAACACCGAGCCAGCAGCTCCAGGGACAGCGCTGG 854
114 LuGlu.....GlnProMetProGluThrProLysTrp 124
   ||| ||| ||| ||| ||| |||
855 CAGAGGTTCTGTCAAGACACCTGCTCCGATTCAGAGAGTCCA..... 898
125 PheHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSerSerLys 141
   ||| ||| ||| ||| ||| |||
899 .....CCTTCAAGAGCAG 912
141 sPheValThrArgValThrProAlaGlnArgGlyLysValProProV 158
   ||| ||| ||| ||| ||| |||
913 AAGCATGTCCATACACAGAGAGTGTGGAGGG..... 949
158 aHisHisSerSerProThrThrGlnIleThrHisArgProArgValSer 174
```

```
950 .....ACCAAGCTCGGTGACAAACAGGGCTAGAGCCAC 985
   :||| ||| ||| ||| ||| |||
175 Arg.....AlaSerSerValAlaGlyAspLysProArg..... 185
   :||| ||| ||| ||| ||| |||
986 AAGGACAGGAGGAGATGACAACCTAGCAAGGCTGATAGCCAAAGGAGGA 1035
186 .....T 186
1036 CATAGAGGGGTCAGGATAGCTTTGATGCACCAAAAGGTCTCTAGGAA 1085
186 hrPheLeuProSerThrThrAlaSerLysIleSerAlaLeuGluGlyLeu 202
   ||| ||| ||| ||| ||| |||
1086 CCAATTGGGCCACGAGCTGTGCTTCAGAAACTTTGGCTCGGAA...ATC 1132
203 LeuLysProGlnThrPro 208
   ||| ||| ||| ||| |||
1133 CTCCCAAGCAACGCCA 1150
```

seq\_name: N\_Geneseq\_36:Q14498

seq\_documentation\_block:

ID\_Q14498 standard; DNA; 1876 BP.

AC\_Q14498;

DT\_10-MAR-1993 (revised)

DE\_30-JAN-1992 (first entry)

KW\_Rabbit; insemination; pregnancy; ss.

OS\_Oryctolagus cuniculus.

FT\_Key Location/Qualifiers

PN\_W09116061-A.

PD\_31-OCT-1991.

PF\_16-APR-1991; U02604.

PA\_16-APR-1990; US-510161.

PA\_(HARD ) HARVARD COLLEGE.

PA\_(SURE-) INST SUISSE RECH EXPER C.

PI\_Kraehenbuhl JP, Weltzin RA, Neutra MR;

DR\_WPI; 91-339549/46.

PI\_PFSDB; R14670.

PT\_Stabilised poly-Ig complex contg. portion of poly-Ig receptor -

PT\_useful in protection against pathogens or against pregnancy

PS\_Disclosure; fig 8; 51 pp; English.

CC\_The sequence was obt'd. from a cDNA clone of allele no. 1 and

CC\_encodes a truncated poly-Ig receptor. The native gene (Mostov et

CC\_al) is mutated to delete the portion encoding the transmembrane and

CC\_intra-cellular domains. The recombinant protein produced by

CC\_expression of the sequence is used as a stabiliser protein with a

CC\_poly-Ig specific for a selected antigen or family of antigens. The

CC\_compn. can be administered directly to the mucosal surfaces of a

CC\_mammal to protect against a pathogen or against insemination. It

CC\_protects against allergens that contact the respiratory or digestive

CC\_mucosal surfaces and protects against pregnancy by cross-linking

CC\_sperm in the vagina.

CC\_See also Q14499.

SQ\_Sequence 1876 BP: 441 A; 525 C; 549 G; 361 T;

alignment\_scores:

Quality: 166.50 Length: 103

Ratio: 2.602 Gaps: 4

Percent Similarity: 62.136 Percent Identity: 42.718

alignment\_block:

US-09-135-238B-2\_COPY\_18\_253 x Q14498 ..

Align seg 1/1 to: Q14498 from: 1 to: 1876

13 GlyLysSerValThrIleLysCysProLeuPro.....GluMe 25

||| ||| ||| ||| ||| ||| |||

238 GCGCACTCGGTGTCATGCTACATGCTACTACCAACACCTCCGTCACCGC 287

25 thisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 42

```

288 GCACAGCGGAGAGTCTGTCGCGCGGAA...GAGGAGAGCGCGCGTCGC 334
42 lyThrValValSerThrThrAsnPhelIleLysAlaGluTyrLysGlyArg 58
335 TGACGCTGCTCG...ACGGCTACAGCTCCAGGAATACTCCGGGAGA 381
59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75
382 GGCAAGCTCACCAGCTCCCTGATAAGGGAGTTTGTGACTGTGA 431
75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 92
432 CCAACTCACCAGAGCACTCAGGAGCTACAAGTGTGGTGGGAGTCA 481
92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 108
482 AC...GGCCGTGGCTGGAGTTCGGTGTCAACGTGTGTCAGCCAGNAAG 528
109 TyrGluPro 111
529 CCAGAGCCT 537

```

seq\_name: N\_Geneseq\_36:T31287

seq\_documentation\_block:

ID T31287 standard; cDNA; 3517 BP.

```

AC T31287;
DT 24-FEB-1997 (first entry)
DE Rabbit poly-immunoglobulin receptor, cDNA.
KW Rabbit; immunoglobulin; receptor; protection protein; mutants;
KW heavy chain; antigen binding domain; protection; pathogen;
KW mucosal; environment; gastrointestinal; passive; immunisation;
KW Guy's 13 antibody; prevention; dental caries; Streptococcus;
KW poly; sorbinus; ss.
OS Oryctolagus cuniculus.
FH Key Location/Qualifiers
FT cds 124..2445
FT /tag= a
PD WO9621012-Al.
PN 11-JUL-1996.
PE 27-DEC-1995; U16889.
PR 30-DEC-1994; US-367395.
PR 04-MAY-1995; US-434000.
PA (PLAN-) PLANT BIOTECHNOLOGY INC.
PA (UNNE-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.
PA (PLAN-) PLANET BIOTECHNOLOGY INC.
PI Hiatt AC, Lehner T, Ma JKC;
DR WPI: 96-333987/33.
DR P-PSDB: W03177.
PT Immunoglobulin and protection protein complex and its prodn. in
PT plants - useful for passive immunisation against mucosal antigens,
PT esp. against S. mutants and S. sorbinus to prevent dental caries
PS Claim 10; Pages 99-102; 152pp; English.
CC The present sequence encodes the rabbit poly-immunoglobulin (Ig)
CC receptor, a portion of which corresp. to residues 1-627, pref.
CC 1-606, or esp. residues 21-43, 1-118, 119-223, 224-332, 333-441,
CC 442-552, 553-606 or 553-627 comprises a protection protein (PP).
CC The Ig of the invention comprises a PP as above in association with
CC an Ig derived heavy chain, having at least a portion of an antigen
CC (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.
CC gastrointestinal, environments, therefore enhancing its
CC effectiveness in passively immunising animals against mucosal
CC pathogens. The Ag binding domain is specifically derived from the
CC Guy's 13 antibody, and the Ig can be used to prevent dental caries
CC by binding, e.g. Streptococcus mutants serotypes c, e and f, or
CC S. sorbinus serotypes d and g.
SQ Sequence 3517 BP; 795 A; 1027 C; 1038 G; 657 T;

```

alignment\_scores:

Quality: 166.50

Ratio: 2.602

Length: 103

Gaps: 4

Percent Similarity: 62.136 Percent Identity: 42.718

alignment\_block:

US-09-135-238B-2\_COPY\_18\_253 x T31287 ..

Align seg 1/1 to: T31287 from: 1 to: 3517

```

13 GlyGlySerValThrIleLysCysProLeuPro.....GluMe 25
238 GCGGACTCGGTGTCCATCATCTACTACCAACAACCTCCGTCACCG 287
25 tHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 42
288 GCACAGCGGAGAGTTCTGGTCCCGGAA...GAGGAGAGCGCGCGTCG 334
42 lyThrValValSerThrThrAsnPhelIleLysAlaGluTyrLysGlyArg 58
335 TGACGCTGCTCG...ACGGCTACAGCTCCAGGAATACTCCGGGAGA 381
59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75
382 GGCAAGCTCACCAGCTCCCTGATAAAGGGAGTTTGTGACTGTGA 431
75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 92
432 CCAACTCACCAGAGCACTCAGGAGCTACAAGTGTGGTGGGAGTCA 481
92 snThrAspArgGlyLysThrGlnLysValThrLeuAsnValHisSerGlu 108
482 AC...GGCCGTGGCTGGAGTTCGGTGTCAACGTGTGTCAGCCAGNAAG 528
109 TyrGluPro 111
529 CCAGAGCCT 537

```

seq\_name: N\_Geneseq\_36:Q14499

seq\_documentation\_block:

ID Q14499 standard; cDNA; 1876 BP.

```

AC Q14499;
DT 30-JAN-1992 (first entry)
DE Clone 2 for truncated poly Ig-receptor.
KW Rabbit; insemination; pregnancy; ss.
OS Oryctolagus cuniculus.
FH Key Location/Qualifiers
FT cds 124..1876
FT WO9116061-A.
PD 31-OCT-1991.
PR 16-APR-1991; U02604.
PR 16-APR-1990; US-510161.
PA (HARD ) HARVARD COLLEGE.
PA (SURE-) INST SUISSE RECH EXPER C.
PI Kraehenbuhl JP, Weltzin RA, Neutra MR;
DR WPI: 91-339549/46.
DR P-PSDB: R14671.
CC Stabilised poly-Ig complex contg. portion of poly-Ig receptor -
CC useful in protection against pathogens or against pregnancy
CC Disclosure; Fig 8; 51 pp; English.
CC The sequence was obt'd. from a cDNA clone of allele no. 2 and
CC encodes a truncated poly-Ig receptor. The native gene (Mostov et
CC al) is mutated to delete the portion encoding the transmembrane and
CC intra-cellular domains. The recombinant protein produced by
CC expression of the sequence is used as a stabiliser protein with a
CC poly-Ig specific for a selected antigen or family of antigens. The
CC compsn. can be administered directly to the mucosal surfaces of a
CC mammal to protect against a pathogen or against insemination. It
CC protects against allergens that contact the respiratory or digestive
CC mucosal surfaces and protects against pregnancy by cross-linking
CC sperm in the vagina.
CC See also Q14498.
SQ Sequence 1876 BP; 357 A; 452 C; 483 G; 306 T;

```

CC produced.. The same method can be used to identify antagonists of PIGR-2  
CC but the presence of an antagonist is indicated by a decrease in  
CC production of the signal. Antibodies against PIGR-2 may be used to  
CC isolate or identify clones expressing PIGR-2.  
SQ Sequence 1137 BP; 252 A; 318 C; 326 G; 241 T;

alignment\_scores:  
Quality: 119.50 Length: 153  
Ratio: 1.532 Gaps: 7  
Percent Similarity: 50.980 Percent Identity: 27.451

alignment\_block:  
US-09-135-238B-2\_COPY\_18\_253 x X28250 ..  
Align seg 1/1 to: X28250 from: 1 to: 1137

13 GlyGlySerValThrIleLysCysProLeuPro.....Glume 25  
||| |||||:|||| ||| |||  
238 GGGACTGGTGTCCATCAGTGTACTACCCACACCTCGGTCAACCG 287  
||| |||||:|||| ||| |||  
25 tHisValArgIleTyrLeuCysArgGluMetAlaGlySerGlyThrCysG 42  
||| ||| :||| ||||| ||| |||  
288 GCACAGCCGGAAGTTCTGTGGCCGGAA...GAGGAGAGCGCGCTGGC 334  
||| |||||:|||| ||| |||

42 lYthrValValSerThrThrAsnPheIleLysAlaGluTyrLysGlyArg 58  
||| |||||:|||| ||| |||  
335 TGACGCTTGCTCG...ACGGCTACAGTCCCGAGGATACTCCGGGAGA 381  
||| |||||:|||| ||| |||  
59 ValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuValGluValTh 75  
||| |||||:|||| ||| |||  
382 GGCAATCTCAGCGACTTCCCTGAAAGGGGGAGTTTGTGGTGTATTAA 431  
||| |||||:|||| ||| |||  
75 rGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyAlaGlyMeta 92  
||| |||||:|||| ||| |||  
432 CCACCTCAGCAGAGCGACTCAGGAGCTAGAGTGTGGGTGGAGTCA 481  
||| |||||:|||| ||| |||

92 sn 92  
||  
482 AC 483

seq\_name: N\_Geneseq\_36:X28250

seq\_documentation\_block:  
ID X28250 standard; cDNA; 1137 BP.  
AC X28250;  
DE 16-JUN-1999 (first entry)  
DT Human PIGR-2 coding sequence.  
KW PIGR-2; human; autoimmune disease; rheumatoid arthritis; psoriasis;  
KW Multiple Sclerosis; Systemic Lupus Erythematosus; diagnosis; therapy;  
KW Inflammatory Bowel Disease; SS.  
OS Homo sapiens.  
PN EP-905237-A2.  
FD 31-MAR-1999.  
PR 21-NOV-1997; US-976293.  
PR 07-AUG-1998; US-056774.  
PR 25-AUG-1997; US-056774.  
PA (SNIK ) SMITHKLINE BEECHAM CORP.  
FI Sweet RW, Truneh A, Wu S;  
DR WPI; 99-192665/17.  
DR P-PSDB; Y05069.  
DR New polypeptides encoding human PIGR-2 useful for treating diseases  
PT such as rheumatoid arthritis and multiple sclerosis  
PS Claim 4; Page 16-17; 23pp; English.

CC This sequence encodes the human PIGR-2 protein of the invention.  
CC Autoimmune diseases involving altered expression or activity of PIGR-2  
CC may include rheumatoid arthritis, Multiple Sclerosis, psoriasis,  
CC Systemic Lupus Erythematosus and Inflammatory Bowel Disease. These  
CC diseases can be diagnosed or susceptibility to them predicted by:  
CC (1) determining whether there is a mutation in the genomic copy of the  
CC gene encoding PIGR-2; or (2) measuring the amount of PIGR-2 in a sample  
CC derived from the patient. Patients deficient in PIGR-2 can be treated by  
CC administering either the PIGR-2 DNA or its complement or an agonist of  
CC PIGR-2 to the patient. Patients with excessive expression or activity of  
CC PIGR-2 can be treated by administering an antagonist of PIGR-2, an  
CC antisense nucleic acid molecule which inhibits the expression of PIGR-2  
CC or administering sufficient PIGR-2 to compete with the endogenous  
CC activity. PIGR-2 can be used to identify its agonists by contacting a  
CC cell expressing PIGR-2 with a candidate compound in the presence of a  
CC signal system and noting the candidate as an agonist if a signal is

8 ValGluGlyGluLeuGlyGlySerValThrIleLysCysProLeuProG1 24  
||| ||| ||| |||:||||| ||| |||  
216 GTGACTGGCACTGCGGGGACTCTCTGACAGTGTGGTGTAGTATGAGAG 265  
||| ||| ||| |||:||||| ||| |||  
24 uMetHis.....ValArgIleTyrLeuCysArgGluMetAlaGlySerG 39  
||| ||| ||| |||:||||| ||| |||  
266 CATGTACAAGGGATATACAACTACTGTGGTCCGAGGACAGTACGACAG 314  
||| ||| ||| |||:||||| ||| |||  
39 lYthrCysGlyThrValValSerThrThrAsnPheIleLysAlaGluTyr 55  
||| ||| ||| |||:||||| ||| |||  
315 ..TCATGTGAGAGCATTTGTGGAGACCAAGGGAGAGAGAGTGGAGAG 362  
||| ||| ||| |||:||||| ||| |||  
56 LysGlyArgValThrLeuLysGlnTyrProArgLysAsnLeuPheLeuVa 72  
||| ||| ||| |||:||||| ||| |||  
363 AATGCCGCGGTGTCATCAGAGACCCCGGAGGCTCTCGCCTTCACTGT 412  
||| ||| ||| |||:||||| ||| |||  
72 lGluValThrGlnLeuThrGluSerAspSerGlyValTyrAlaCysGlyA 89  
||| ||| ||| |||:||||| ||| |||  
413 GACCATGAGAACCTCAATGAAGATGATGCTGGATCTTACTGTGTC.... 458  
||| ||| ||| |||:||||| ||| |||  
89 laGlyMetAsnThrAspArgGlyLysThrGlnLysValThrLeuAsnVal 105  
||| ||| ||| |||:||||| ||| |||  
459 .....AAAAATTCAGACAGTGTGGTGC..... 479  
||| ||| ||| |||:||||| ||| |||  
106 HisSerGluTyrGluProSerTrpGluGluGlnPrometProGluThrPr 122  
||| ||| ||| |||:||||| ||| |||  
480 .....CTGGATTTCATGTCACGCGATCCCTCG..... 506  
||| ||| ||| |||:||||| ||| |||  
122 oLysTrpPheHisLeuProTyrLeuPheGlnMetProAlaTyrAlaSerS 139  
||| ||| ||| |||:||||| ||| |||  
507 .....GACCTGGTTAGGGTGTATGTTTCCCGAGCAATTACACCC 546  
||| ||| ||| |||:||||| ||| |||  
139 erSerLysPheValThrArgValThrThrProAlaGlnArgGlyLysVal 155  
||| ||| ||| |||:||||| ||| |||  
547 CA.....AGGAGGACCACACATCCAGCC.....ACA 572  
||| ||| ||| |||:||||| ||| |||  
156 ProProVal 158  
||| ||| ||| |||:||||| ||| |||  
573 CCTCCCATC 581





